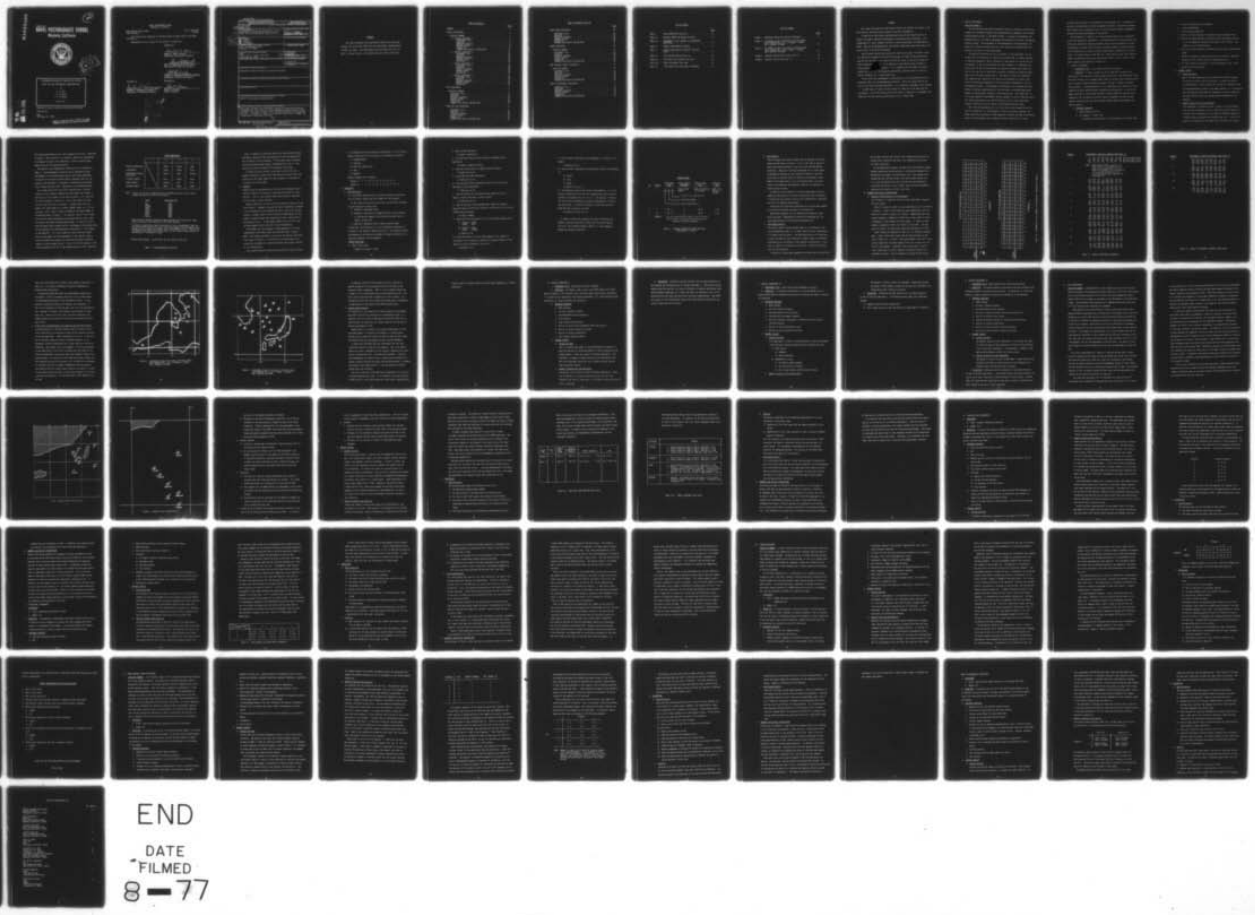
AD-A042 305 NAVAL POSTGRADUATE SCHOOL MONTEREY CALIF F/G 17/2
EXPERIMENTAL DESIGNS AND ANALYSES FOR INITIAL ACCAT TEST BED EX--ETC(U)
APR 77 D R BARR, G K POOCK, F R RICHARDS
UNCLASSIFIED NPS55-77-21 NL

| OF |
ADA042305



END

DATE FILMED
8-77

(2)

NPS55-77-21

NAVAL POSTGRADUATE SCHOOL

Monterey, California

EXPERIMENTAL DESIGNS AND ANALYSES FOR INITIAL

ACCAT TEST BED EXPERIMENTAL DEMONSTRATIONS

by

D. R. Barr
G. K. Poock
F. R. Richards

April 1977

Prepared for:

NOSC
San Diego, CA 92152

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Isham Linder
Superintendent

Jack R. Borsting
Provost

This research was supported by the Naval Ocean Sciences Center, San Diego, California.

Prepared by:

Donald R Barr

DONALD R. BARR, Professor
Department of Operations Research

Gary K Pooch

GARY K. POOCH, Associate Professor
Department of Operations Research

F R Richards

FRANCIS R. RICHARDS, Associate Professor
Department of Operations Research

Reviewed by:

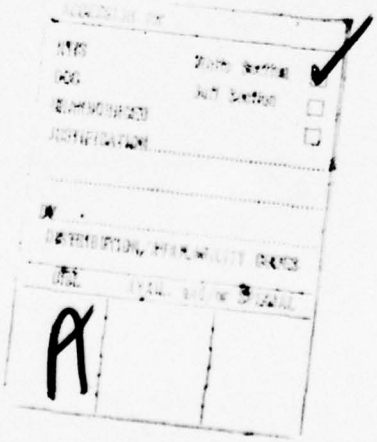
Michael G Sovereign

MICHAEL G. SOVEREIGN, Chairman
Department of Operations Research

Released by:

Robert R Fossum

ROBERT R. FOSSUM
Dean of Research



Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS55-77-21
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Experimental Designs and Analyses for Initial ACCAT Test Bed Experimental Demonstrations.

5. TYPE OF REPORT & PERIOD COVERED: Technical Report

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): D. R. Barr, K. Poock, R. Richards

8. CONTRACT OR GRANT NUMBER(s):

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School, Monterey, CA 93940

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS: NOSC, ATTN: CDR C. A. Rose, Code 1020, San Diego, CA 92152

12. REPORT DATE: April 1977

13. NUMBER OF PAGES: 77 (12) 83p.

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

ACCAT; C^2 Experiments, Command Control

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report documents the experimental designs and associated analyses recommended for the initial test bed experimental demonstrations at NOSC, San Diego. The areas in which initial demonstrations are to be conducted include: 1) displays; 2) TECA; 3) man-machine interface with LADDER; and 4) man-machine interface with RITA.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251450

ABSTRACT

This report documents the experimental designs and associated analyses for the initial ACCAT test bed experimental demonstrations at NOSC, San Diego. The experimental team performed this work for CDR C. A. Rose, Director of the ACCAT project at NOSC (Code 1020).

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This report describes the experimental designs and analyses to be used in the four specific areas that were assigned to the NPS investigators.

In addition to developing experimentation plans to meet specific goals of ACCAT, this report represents development of experimentation technology. This technology will itself undergo "testing" as future experiments are undertaken at ACCAT. Much of the methodology for the specific experiments described herein will prove useful in follow-on experiments.

Due to normal problems associated with starting a project like ACCAT, there are limitations imposed on the real-world aspects of these initial experiments. Looking into the future, the results of what one learns in the early stages will affect subsequent research. It is necessary to get operationally oriented in the experiment as soon as possible. It is important for the ACCAT test bed to develop the capability for producing experimental resources such as scenarios, special software, etc., as the needs arise.

Pilot trials are recommended in some sections to allow determination of appropriate levels of the experimentation factors. While this phase need not be part of the formal recorded experiments, it is important to document their outcome.

The NPS team will need at least 30 days for analysis of the data from the experiments. As of this writing, the final report is expected on 1 September 1977. Therefore, the data should be delivered to us by 1 August 1977.

I. DISPLAY EXPERIMENTS

Executive Summary.

This section describes those experiments which are planned to evaluate the transfer of information between the software/hardware components and the human operator/decision maker in the ACCAT testbed. A variety of experiments are proposed which will provide a comprehensive evaluation of the total information transfer system. The development of the methodologies and experimental concepts will provide an output which will form the basis of methodologies and concepts to be used in future experiments.

It would not be feasible to try to evaluate, in one experiment, all the display variables which are of concern to the ACCAT team. A modular approach is therefore proposed in which a sequence of experiments will be carried out. For example, in the first experiment the total economics of a GENISCO color display will be one of the variables of interest (economics here refers not only to cost but also to the effect on operator information processing times, usability of the total system, etc.). The results of the first experiment with regard to this question, will then form a guideline for colors to be used in the display presentation in the next module (experiment) of the experimentation process. The sequence of experiments will be phased such that results of previous experiments will form the basis of upcoming experiments needed to evaluate other variables in the display area. The longer goal is to tie the sequential experimental results together to form an effective basis for reaching design decisions (specifications) on how to most effectively transfer information from the system to the operator/decision maker.

Data from the experiments will be in the form of both subjective and objective information, i.e. measurements of times, accuracies, opinions, etc. Proper statistical analyses of these dependent variables can then be performed which will give answers such as whether different conditions of a given

variable affect transfer of information to the operator, etc. In addition to gaining information about various independent variables, interactions between the variables can be assessed as well as the correlation between independent variables and human performance in the system.

The following is a partial list of the variables which will be examined during the sequence of experiments: low resolution color displays versus high resolution black/white displays, mercator versus polar map projection, types of geographic locations and situations represented therein; function buttons versus keyboard entry; where to present alpha numeric information of various types - on status boards, on the CRT, etc.; NTDS symbols versus other designs; method of presenting ship tracks (bearing and speed); size of display screen really needed; zoom capability on maps; etc.

A. DISPLAY EXPERIMENT I

1. EXPERIMENT TITLE: Color Combination-Naval Situation

2. OBJECTIVE: To make a comparison of the usability, economics and feasibility of various color display combinations versus the conventional black and white display, and to compare these under 24 various naval situations which vary in type of threat presented and number of enemy, friendly and neutral forces present. A second objective in this experiment is to evaluate the presentation of newly acquired enemy information by displaying it in one of two modes, blinking or non-blinking. Measures of effectiveness to evaluate these objectives will be operator proximity threat assessment time and accuracy, operator detection time for newly presented enemy symbols and operator subjective opinions.

3. RESOURCES REQUIRED:

a. ACCAT testbed facility

b. 20 subjects - 3 hours each

(2 subjects can participate in the experiment at the same time)

c. 24 Naval situations to be displayed

d. 1 (one) experimenter

e. 1 color display (GENISCO)

f. Ability to record operator's proximity threat assessment time

g. Ability to record operator's proximity threat assessment accuracy (i.e. the number of platforms he believes to be a threat. See details in 4.b. which follows)

h. Ability to record operator's detection time of newly inserted enemy symbol

i. Record of operator's subjective opinions to a series of questions presented on the display during his debriefing period

j. All data should be provided on 9 track IBM magnetic tape or IBM cards for final evaluation

4. GENERAL CONTEXT

a. Concept and Need.

One of the new technologies to be examined in the ACCAT testbed is the use of a multi-color CRT display. In order to evaluate its effectiveness in facilitating the transfer of information from the software/hardware system to the human operator, it is imperative that the displays experiments be undertaken in order to determine the most effective means for presenting the information to the operator.

b. General Situation for the Experiment.

This experiment is designed so that a subject will be presented with 24 naval situations, one at a time, and in each situation, a different color combination will be used to represent enemey and friendly forces, longitude and latitude lines, etc. A trial will consist of displaying to the subject one situation with a given

color combination. The operator's main task will be to make a proximity threat assessment of the situation. The experimental conceptual design is shown below in Figure 1. The 24 naval situations displayed are part of the resources required and will be developed by the ACCAT project team. (See details in Section 6.b. - Comments and Special Instructions.)

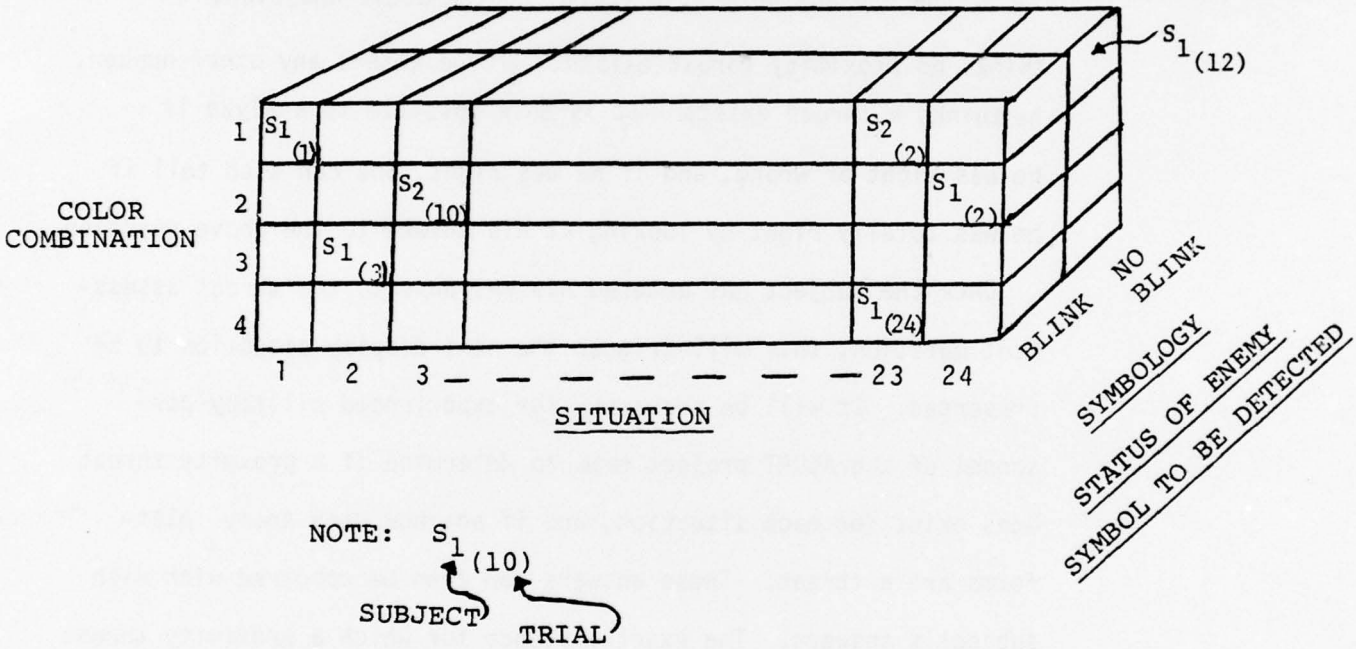


Figure 1. CONCEPTUAL DESIGN

Accordingly then, a given subject will be presented a display of a naval situation, he will make a proximity threat assessment of the situation and when finished, type an F (for finished) into the keyboard associated with the GENISCO display. When the situation is presented on the display screen, a clock will start running and subsequently stop when the operator depresses the F key, thus obtaining a measure of assessment time. When the F key is

depressed, it will not only stop the clock but also cause the following question to appear on the display screen (or a side screen):

"If you believe there was no threat in the display you just observed, type a zero (0) on the keyboard. If you believe there was a threat, enter on the keyboard the number of enemy platforms which you considered to be a threat."

(Subject types in a number.)

If the subject enters 0 (zero) to the above question, he thinks no proximity threat exists. If he enters any other number, he thinks a threat exists. It is then possible to analyze if he was right or wrong, and if he was right, one can also tell if he was totally right by looking at his answer to the above question.

Once the subject has entered his response to the threat assessment question, this will trigger the next display situation to be presented. It will be necessary for experienced military personnel of the ACCAT project team to determine if a proxmity threat does exist for each situation, and if so, how many enemy platforms are a threat. These answers can then be compared with each subject's answers. The exact distance for which a proximity threat exists will be determined in pilot trials described in Section 6.b.

It is anticipated the proximity threat assessment will take around 30 seconds to 1 minute. Thus, in a given experimental session, a subject will receive eight (8) displays before having a rest period. Consequently, it is planned to run each subject in 3 experimental sessions of eight trials each (see typical schedule for subjects which is described later).

To summarize the foregoing, subject 2 for example, will enter the experimental area, be indoctrinated and receive some practice,

and then be presented with his first sequence of 8 trials. Referring to Figure 1, the second trial for subject 2 might be a presentation of situation 23 under color combination 1 with a blinking enemy symbol which will be explained shortly.

The specific color combinations referred to are illustrated in TABLE I. The monochromatic condition can be obtained by using the GENISCO in a black/white mode with no color being generated. In addition to the main proximity threat assessment task for the observer, the experiment is designed to investigate the presentation of newly acquired enemy information in two different modes.

At a random time within the first 5 seconds after a display situation (trial) begins, another new enemy symbol will appear on the display. It will appear on the display screen 1) as a new symbol being constantly displayed or 2) as a new symbol but which blinks on and off every .25 seconds. In other words, the new symbol will simply appear as a steady lighted symbol or it will be blinking. This is what is meant by STATUS OF THE ENEMY SYMBOL TO BE DETECTED shown in Figure 1. The subjects' task will be to make the proximity threat assessment as described earlier, and in addition detect the newly presented enemy symbol. A clock will start when the enemy symbol appears and will be stopped by the subject when he has detected it by the depression of the letter D (for detect) on the keyboard. It is also planned to simply have the subject point to the symbol and the experimenter in the room will keep a record of whether the subject had identified the newly presented symbol or simply thought he had. This will be fairly obvious in the blinking condition but not in the steady state (no blink) condition.

	COLOR COMBINATION			
	1	2	3	4
Display Background	MONOCHROMATIC	Black	Black	Aqua
Land Masses		Green	Green	Blue
Longitude-Latitude Grid Lines		Green	Blue	Purple
Friendly Symbols		Blue	Green	Blue
Enemy Symbols		Orange	Red	Red
Unknown Symbols		Amber	Amber	Amber

NOTE: If many colors can be generated from the GENISCO display, then we would suggest the specific colors listed below:

Color	Angstroms (A°)
Red	6420
Orange	6100
Amber	5960
Yellow	5820
Aqua	5560
Green	5150
Blue	4760
Purple	4300

These specific spectral colors are chosen because past research has shown them to be equally discriminable to the human observer.

If Angstrom comparisons cannot be obtained or set on the GENISCO, then the next best choice would be to select color book numbers from the Munsell color system and try to match the display color to the Munsell color. The following Munsell book numbers are colors which are also equally discriminable to the observer in the Munsell system:

Munsell Book Numbers: 3R, 9R, 9YR, 1GY, 3G, 7BG, 9B, 9PB, 3RP

TABLE I. COLOR COMBINATION SPECIFICS

Thus, in addition to time and accuracy for the proximity threat assessment, detection time and accuracy will also be obtained for each situation (trial) presented. If the subject does not detect the newly presented enemy symbol, the depression of the F key should also stop the running of the detection time clock.

It should be fairly obvious by now that a given trial has a subject looking at a given naval situation with a given color combination and the enemy symbol for that situation is presented in a blinking or no-blink mode.

c. Subjects

Subjects must be representative of the real world command control officer; hopefully the subjects would actually be some of those officers. That is, the subjects should have previous experience with the types of tasks required.

Subjects should not be color blind (for proper evaluation of the color display). However, since we are really interested in whether or not the subject can distinguish the colors used on the current display, a simpler check will be to have the experimenter ask the subject to describe each color during the initial indoctrination and practice trials. If the subject identifies red as red, etc., the experimenter should note this or any discrepancies.

Each subject will be assigned a random sequence of 24 trials (experimental conditions) which are described later in detail. (NOTE: Each subject performs in only 24 of the possible 192 conditions shown in Figure 1.) The 24 trials for each subject will actually be presented in 3 sessions of eight trials, each followed by a 15 minute break.

Each subject should be asked to be available for about 3 hours.

It is proposed to run two subjects simultaneously in the following sequence where each of the following is 15 minutes in duration:

I = Indoctrination

P = Practice

W = Work for eight trials

R = Rest

D = Debrief

Typical Schedule for 2 Subjects:

Subject 1: I - P - R - W - R - W - R - W - R - D

Subject 2: I - P - R - W - R - W - R - W - R - D

5. EVALUATION

a. Data Collection

It is extremely important that all data be correctly associated with the proper subject and trial number for that subject.

As described in the previous sections, the following variables would be measured for each trial:

1) Operator's assessment time

2) Operator's entering of a number after he's done assessing to indicate the number of platforms he believes to be a threat in that trial

3) Operator's detection time of newly acquired symbol

In addition, the following is a list of subjective questions to be asked in the subject's debriefing period. The questions should be presented on the display and subject's answers recorded by his entering responses into the keyboard.

DEBRIEF QUESTIONS:

1. What is your rank?

(subject (S) types in rank)

2. What is your specialty?

(S types in specialty)

3. How many years have you been involved in command control operations?

(S types in number of years)

4. Have you ever served as a command control officer?

(S types yes or no)

5. Do you have any NTDS experience?

(S types yes or no)

6. Have you ever had any experience similar to the tasks you performed in this experiment?

(S types yes or no)

7. Did the blinking of the unknown enemy symbol help you to detect it easier than when it didn't blink?

(S types yes or no)

8. What do you feel is the maximum total number of friendly, enemy and neutral platforms an operator could handle on the display without being confused?

(S types a number)

9. Which colors do you prefer, A or B, for friendly-enemy colors?

A) Friendly - Green
Enemy - Red

B) Friendly - Blue
Enemy - Orange

(S types A or B)

10. Would you prefer to use (A) NTDS symbols or (B) symbols in the shape of an airplane to indicate an airplane, symbols in the shape of a ship to indicate a ship, etc.?

(S enters A or B)

11. Do you prefer land masses to be displayed in A) Blue or B) Green?

(S enters A or B)

12. Do you prefer longitude and latitude grid lines to be displayed in:

A) Purple

B) Blue

C) Green

(S enters A or B or C)

13. This question requires one slide to be presented. It is presented 4 times and each of the color combinations used in the experiment is presented once (A, B, C, D). The subject is asked to tell which color combination he prefers (monochromatic is one of the conditions). S should be allowed to switch back and forth among the four to make up his mind.

(S enters A or B or C or D)

In order to effectively assemble the above information and properly associate each piece of data with the proper subject and trial, the following example, TABLE II, is one suggested method for formatting the data.

EXAMPLE FORMAT

S#	TRIAL#	CONDITIONS CODE	THREAT ASSESS. RESPONSE	THREAT ASSESS. TIME	DETECTION TIME
1	1	1 , 1 , 1	Some Integer number	Some # to nearest .01 seconds	Some # to nearest .01 seconds
			Blink (1) or No Blink (2)		
			Situation Number		
			Color Combination #		
1	2	2, 24 , 1	3	41.24	5.73
1	12	1, 24 , 2	0	38.76	15.22
1	DEBRIEF	Lt, CC, 5, NO, YES, NO, YES, 12, A, B, B, B, C			
		Answers to debrief questions			

TABLE II. POSSIBLE FORMATTING SCHEME FOR DATA
(Referenced to Figure 1)

b. Data Analysis

Times to analyze and times to detect will be analyzed in the conceptual design shown in Figure 1. This is the type of design in which each subject performs in only 24 of the 192 experimental conditions. Analysis of variance techniques with a linear model incorporating selected two way interaction terms will be used. Since times to perform tasks are typically distributed as some member of the Gamma family, it is anticipated that a log transform of the assessment and detection times will be necessary to stabilize the variance.

In addition to the analysis of variance techniques, multiple comparison techniques will be used to determine exact differences in significant main effects such as color combinations, situations and status of detected enemy symbol.

Correlation and regression techniques will be used where needed for association between variables or for prediction.

Debriefing responses will be tabulated and statistical summaries provided. Nonparametric correlation techniques will be used where possible on the debriefing response data.

c. Anticipated Results

The results should indicate whether there is a difference in the color combinations used, i.e. it might show that color combination 3 is better than the others. The same should occur for situations, in that the analysis could indicate if certain types of situations presented cause a difference in the operator's performance. Likewise, if there is a difference in the blink - no-blink conditions, this will be indicated.

In addition, interactions between the variables may be significant

which might indicate that certain color combinations work best for certain naval situations and other color combinations work best for other situations.

For the debriefing response, one may find a relationship between operator experience and performance, a relationship between color preferences for symbols and performance, etc. The subjective responses may also provide insight into future symbol design, etc. It is also anticipated that the situations may be subjectively classified into three or more clutter levels and then performance analyzed with respect to clutter level also.

6. COMMENTS AND SPECIAL INSTRUCTIONS

a. Subjects' Experimentation Trial Sequences

The following is a detailed outline of each experimental sequence for each subject.

Table III assigns a number to every experimental cell (condition) in Figure 1. Table IV then shows the exact experimental cells of Table III in which a subject will perform. Remembering that each subject performs 3 sequences of eight trials each, Table IV gives 3 sequences of cell numbers for each subject. In Table IV, each subject's sequential trials are listed in exact order. The subject will take a 15 minute break after each sequence of 8 trials.

Table V shows the sequences of Table IV in a different arrangement. Table V shows each subject who has been assigned to each cell of Table III. In Table V, the upper number in the cell is the subject number and the lower number in the cell is the trial number for that subject. Therefore, Table IV and Table V can be cross referenced to verify a subject's sequential order of experimental testing. Using 16 subjects providing 24 data points

THIS MATRIX FOR BLINK CONDITION

COLOR COMBINATION	SITUATION #																							
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
2	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
3	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72
4	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96

THIS MATRIX FOR NO-BLINK CONDITION

COLOR COMBINATION	SITUATION #																							
1	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
2	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144
3	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
4	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192

TABLE III. ASSIGNMENT OF CELL NUMBERS TO EXPERIMENTAL CONDITIONS OF FIGURE 1.

SUBJECT	EXPERIMENTAL CONDITION SEQUENCE FROM TABLE III	
1	1, 60, 114, 158, 143, 76, 8, 106	←1st eight trials
	187, 70, 21, 53, 174, 122, 137, 181	←2nd eight trials
	96, 44, 40, 155, 33, 99, 151, 87	←3rd eight trials
	NOTE: Above means subject 1 has his 1st trial in cell 1 of TABLE III (i.e. under color combination #1, situation #1, blinking). His second trial is in cell 60 of TABLE III (i.e. color combination #3, situation 12, blinking) etc.	
2	91, 130, 54, 11, 190, 121, 88, 167	
	116, 81, 17, 148, 109, 66, 24, 103	
	32, 45, 180, 134, 159, 170, 51, 29	
3	185, 142, 55, 14, 43, 104, 192, 83	
	39, 164, 125, 3, 112, 141, 177, 58	
	18, 100, 61, 95, 156, 30, 146, 73	
4	151, 79, 62, 183, 22, 36, 128, 154	
	97, 6, 67, 189, 171, 85, 26, 140	
	120, 57, 47, 107, 16, 90, 149, 124	
5	10, 64, 179, 117, 173, 78, 20, 75	
	132, 162, 71, 52, 169, 46, 7, 113	
	153, 37, 129, 115, 135, 74, 168, 38	
6	13, 68, 94, 150, 182, 23, 144, 101	
	49, 42, 175, 80, 28, 123, 165, 65	
	35, 111, 163, 84, 9, 98, 136, 178	
7	82, 48, 86, 172, 119, 131, 19, 5	
	133, 118, 188, 25, 127, 56, 63, 186	
	102, 147, 153, 93, 41, 160, 12, 50	
8	157, 139, 69, 27, 145, 176, 72, 4	
	184, 191, 126, 15, 34, 77, 105, 108	
	92, 59, 31, 110, 2, 138, 166, 89	
9	110, 9, 66, 128, 140, 24, 46, 100	
	121, 54, 15, 155, 43, 95, 189, 160	
	173, 26, 75, 79, 106, 180, 61, 161	
10	83, 41, 192, 171, 102, 57, 2, 45	
	135, 157, 80, 12, 163, 118, 71, 62	
	29, 76, 145, 175, 138, 20, 112, 130	
11	89, 21, 103, 148, 169, 94, 34, 37	
	5, 167, 120, 59, 88, 108, 153, 186	
	50, 8, 123, 174, 134, 68, 139, 39	

TABLE IV. SUBJECT EXPERIMENTAL SEQUENCES

SUBJECT	EXPERIMENTAL CONDITION SEQUENCE FROM TABLE III
12	183, 115, 104, 86, 178, 17, 31, 99
	92, 166, 144, 47, 150, 181, 64, 52
	125, 93, 49, 122, 11, 156, 18, 33
13	151, 1, 91, 142, 185, 23, 165, 176
	105, 146, 44, 72, 82, 132, 51, 85
	111, 184, 28, 42, 14, 53, 126, 107
14	188, 56, 101, 147, 90, 22, 168, 10
	87, 191, 84, 170, 158, 25, 6, 65
	117, 40, 124, 109, 67, 35, 129, 127
15	152, 63, 55, 97, 162, 119, 48, 78
	190, 16, 60, 74, 154, 19, 38, 179
	116, 137, 77, 4, 27, 177, 133, 141
16	96, 131, 149, 30, 81, 58, 113, 73
	98, 172, 114, 70, 187, 32, 182, 69
	7, 36, 164, 136, 3, 143, 13, 159

TABLE IV. SUBJECT EXPERIMENTAL SEQUENCES (CONTINUED)

THIS MATRIX FOR BLINK CONDITION

COLOR COMBINATION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1 1	8 21	3 12	8 8	7 8	4 10	5 15	1 7	6 21	5 1	2 4	7 23	6 1	3 4	8 12	4 21	2 11	3 17	7 7	5 7	1 11	4 5	6 6	2 15
2	7 12	4 15	8 4	6 13	2 24	3 22	8 19	2 17	1 21	8 13	6 17	4 6	5 18	5 24	3 9	1 19	7 21	6 10	3 5	1 18	2 18	5 14	4 19	7 2
3	6 9	7 24	2 23	5 12	1 12	2 3	3 3	7 14	4 18	3 16	8 18	1 2	3 19	4 3	7 15	5 2	6 16	2 14	4 11	6 2	8 3	1 10	5 11	8 7
4	3 24	5 22	5 8	1 6	8 14	5 6	4 2	6 12	2 10	7 1	3 8	6 20	4 14	7 3	1 24	2 7	8 24	4 22	2 1	8 17	7 20	6 3	3 20	1 17

SITUATION #

THIS MATRIX FOR NO-BLINK CONDITION

COLOR COMBINATION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	4 9	6 22	1 22	3 18	6 8	7 17	2 16	3 6	8 15	1 8	4 20	8 16	2 13	8 20	6 18	3 13	5 16	1 3	5 20	2 9	5 4	7 10	7 5	4 17
2	2 6	1 14	6 14	4 24	3 11	8 11	7 13	4 7	5 19	2 2	7 6	5 9	7 9	2 20	5 21	6 23	1 15	8 22	8 2	4 16	3 14	3 2	1 5	6 7
3	8 5	3 23	7 18	2 12	4 23	6 4	1 23	5 17	7 19	4 8	1 20	3 21	8 1	1 4	2 21	7 22	4 1	5 10	6 19	3 10	6 15	8 23	2 8	5 23
4	5 13	2 22	4 13	7 4	5 5	1 13	6 11	8 6	3 15	6 24	5 3	2 19	1 16	6 5	4 4	8 9	3 1	7 16	1 9	7 11	4 12	2 5	8 10	3 7

SITUATION #

TABLE V. TESTING CONDITIONS FOR SUBJECTS (1-8) FOR EACH TRIAL. UPPER NUMBER IN CELL IS SUBJECT NUMBER, LOWER NUMBER IS TRIAL NUMBER FOR THAT SUBJECT.

THIS MATRIX FOR BLINK CONDITION

COLOR COMBINATION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	13 2	10 7	16 21	15 20	11 9	14 15	16 17	11 18	9 2	14 8	12 21	10 12	16 23	13 21	9 11	15 10	12 6	12 23	15 14	10 22	11 2	14 6	13 6	9 6
2	14 14	9 18	15 21	13 19	10 17	16 4	12 7	16 14	12 24	11 7	14 22	16 18	11 8	15 15	11 24	14 18	10 2	13 20	9 13	13 11	10 8	9 7	12 12	15 7
3	12 19	11 17	13 15	12 16	13 22	9 10	15 3	14 2	10 6	16 6	11 12	15 11	9 23	10 16	15 2	12 15	14 16	9 3	14 21	11 22	16 16	16 12	10 15	13 12
4	16 8	15 12	9 19	10 18	15 19	15 8	9 20	10 11	16 5	13 13	10 1	14 11	13 16	12 4	14 9	11 13	11 1	14 5	13 3	12 9	12 18	11 6	9 14	16 1

SITUATION #

THIS MATRIX FOR NO-BLINK CONDITION

COLOR COMBINATION	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	15 4	16 9	12 8	9 8	14 3	10 5	11 3	12 3	13 9	9 21	13 24	11 14	14 20	9 1	13 17	10 23	16 7	16 11	12 2	15 17	14 17	10 14	15 6	11 11
2	9 9	12 20	11 19	14 19	12 17	13 23	14 24	9 4	14 23	10 24	16 2	13 14	15 23	11 21	10 9	16 20	15 18	10 21	11 23	9 5	15 24	13 4	16 22	12 11
3	10 19	13 10	14 4	11 4	16 3	12 13	13 1	15 1	11 15	15 13	9 12	12 22	10 10	14 13	16 24	9 16	9 24	15 5	10 13	16 19	13 7	12 10	11 10	14 7
4	11 5	14 12	10 4	16 10	9 17	11 20	10 20	13 8	15 22	12 5	15 16	9 22	12 14	16 15	12 1	13 18	13 5	11 16	16 13	14 1	9 15	15 9	14 10	10 3

SITUATION #

TABLE V. CONTINUED FOR SUBJECTS (9-16).

each, the final design will include 2 data points in each cell of Table III. All subject assignments have been randomized and balanced where appropriate.

It should be noted that there are experimental sequences for 16 subjects. These 16 sequences must all be run in these orders to allow proper completion of the experimental design. In the section on resources required, 20 subjects were requested. If the first 16 subjects all provide data, one could stop data gathering. However, if subject 3 for example, had breakdown of equipment, etc., during his trials, then it would be necessary that another subject be assigned to the exact same sequence which subject 3 had.

b. Pilot Trials and Development of Situations and Proximity Threat

As mentioned earlier, the ACCAT project team will develop the 24 naval situations to be displayed as part of the resources required to run the experiment. These situations will vary in 1) the number of friendly, enemy and neutral platforms present, 2) the geographic location represented and 3) the location of the given platforms. See Figures 2 and 3 as examples. It will be necessary to run some initial pilot trials to ensure that the level of difficulty represented on the plates spans the threshold range of the human operator, i.e. the pilot trials need to be done to make sure the situations are not all too easy or all too hard for an operator. If such were the case, no useful information would be gained as the subjects would all assess every situation very easily, or none of the subjects would be able to assess the situation if all situations were too difficult. In this experiment, NTDS symbology will be used.

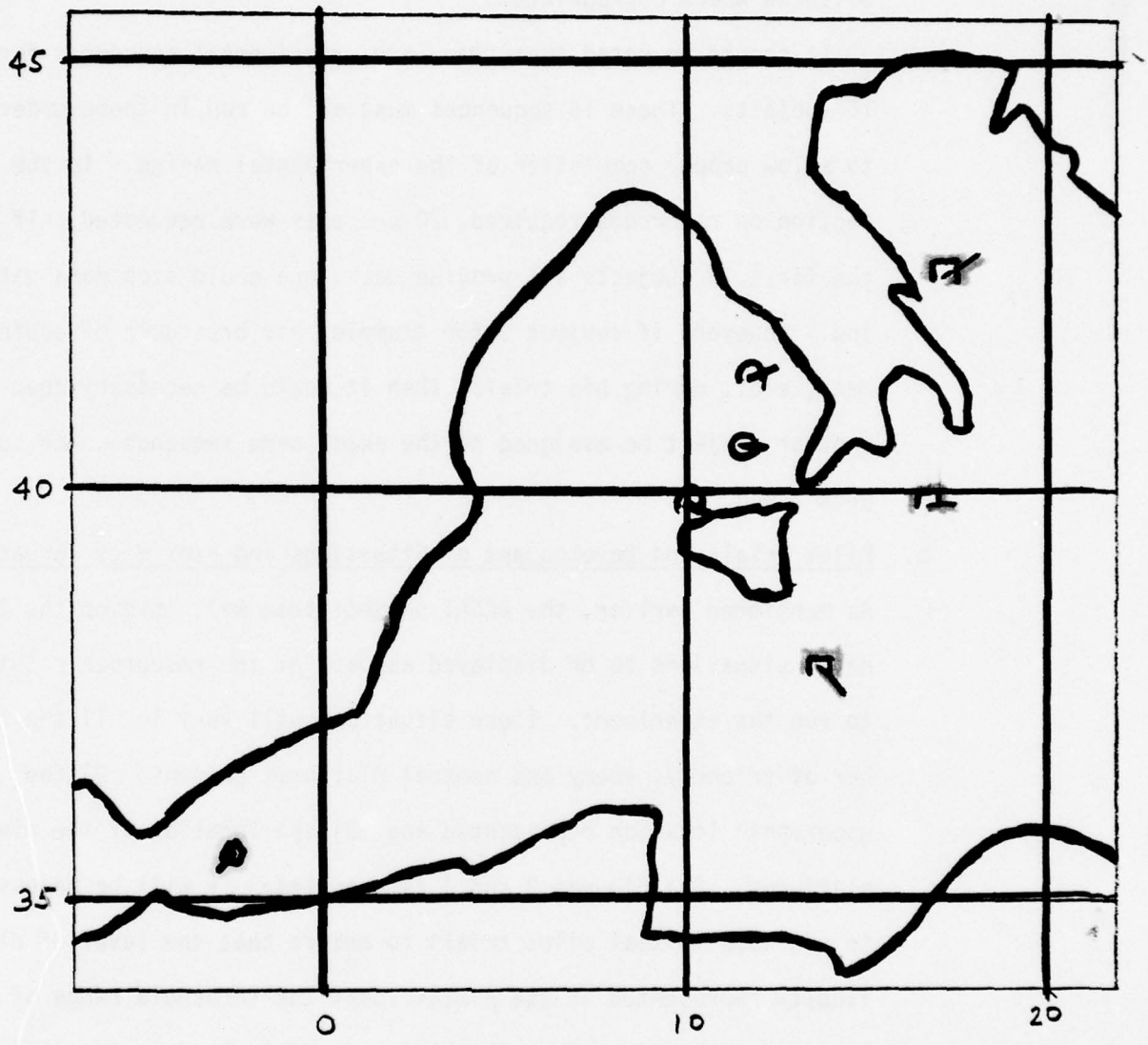


FIGURE 2. AN EXAMPLE OF ONE OF THE NAVAL SITUATIONS UNDER COLOR COMBINATION 2 WITH 3 FRIENDLY, 3 ENEMY AND 1 NEUTRAL PLATFORM.

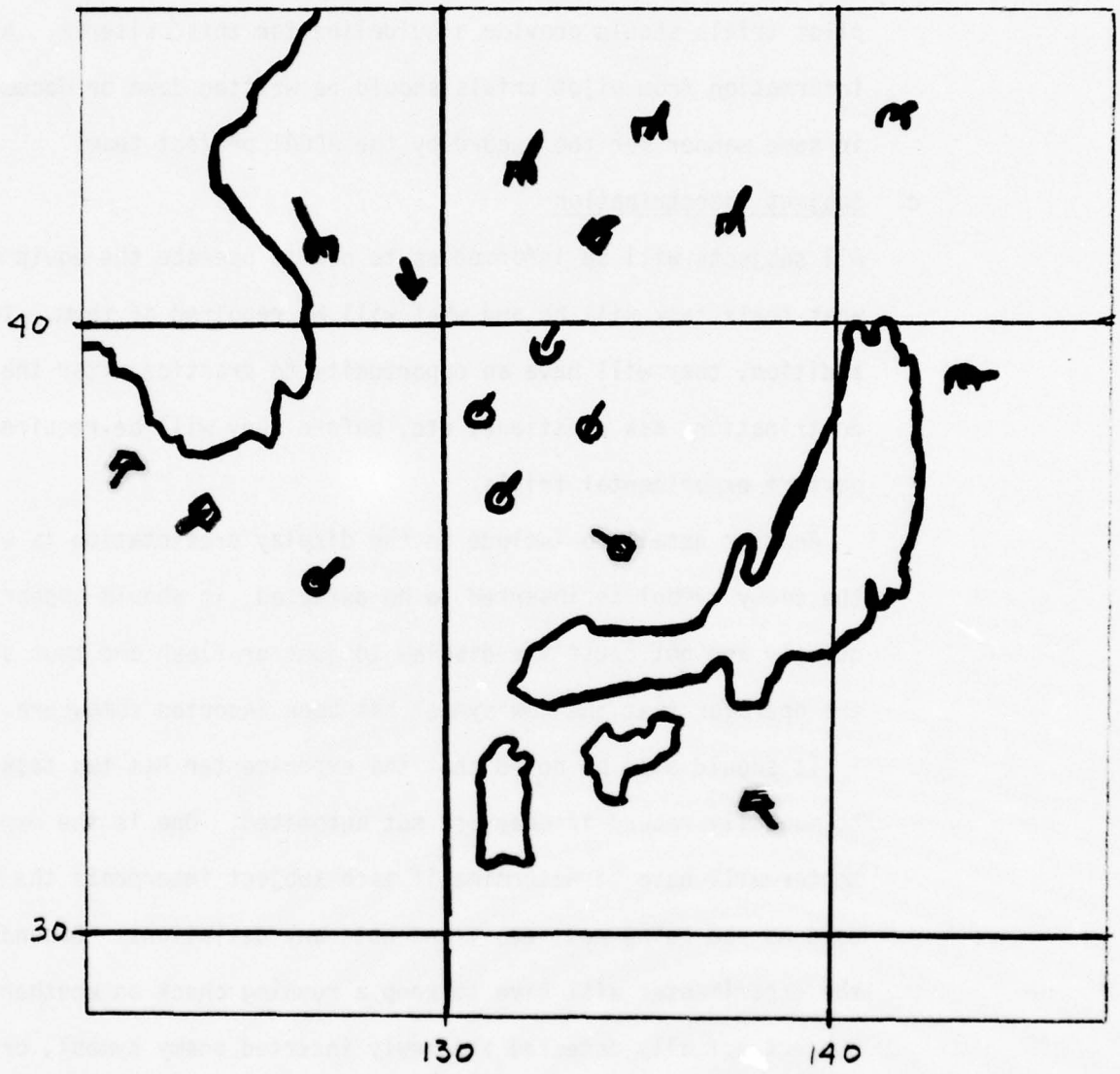


FIGURE 3. AN EXAMPLE OF ONE OF THE NAVAL SITUATIONS UNDER COLOR COMBINATION 3, WITH 7 ENEMY, 7 FRIENDLY AND 3 NEUTRAL PLATFORMS.

In addition, the pilot trials need to be run to provide a general guideline for the distance criteria to be used in the proximity threat assessment. That is, will an enemy platform be a threat if it is within 300 miles or within 600 miles, etc. The pilot trials should provide a guideline for this criteria. All information from pilot trials should be written down or documented in some manner for the record by the ACCAT project team.

c. Subject Indoctrination

All subjects will be informed as to how to operate the equipment, what their task will be and what will be required of them. In addition, they will have an opportunity to practice after the indoctrination, ask questions, etc. before they will be required to perform experimental trials.

Another detail to include in the display presentation is when the enemy symbol is inserted to be detected, it should appear quietly and not cause the display to jump or flash and thus signal the operator that the new symbol has been inserted somewhere.

It should also be noted that the experimenter has two tasks to manually record if they are not automated. One is the experimenter will have to determine if each subject interprets the colors used as red being red, etc., and note any deviations. Secondly, the experimenter will have to keep a running check on whether the subject actually detected the newly inserted enemy symbol, or whether the subject simply depressed the D key and pointed to another symbol which was incorrect.

The pilot trials will also help verify separations to be used on latitude-longitude grid lines. It is presently felt that displays of 600 miles on a side should show grid lines every 5 degrees apart.

Large or smaller displays should show grid lines separated in a similar proportion.

B. DISPLAY EXPERIMENT II

1. EXPERIMENT TITLE: Resolution and Type of Symbols

2. OBJECTIVE: To compare, under a given task, NTDS symbols with other types of symbols in 24 different naval situations under three display conditions: 1) the best color combination from Display Experiment I, 2) Black and White high resolution, 3) Black and White low resolution.

3. RESOURCES REQUIRED:

a. ACCAT Testbed

b. 24 naval situation displays

c. One experimenter and 20 operators

d. Color Display

e. High resolution B/W display

f. Ability to record task performance times and accuracy

g. Ability to record subjective opinions

h. Results of Display Experiment I

i. Set of newly designed symbols

4. GENERAL CONTEXT

a. Concept and Need

This is a second step needed in the experimentation sequence to show the validity of using NTDS symbols or other appropriately designed symbols. Given the results of Display Experiment I, this experiment is needed to compare the best color combination from the first experiment versus the use of black and white low and high resolution displays.

b. General Situation for the Experiment

The design is very similar to that of Display Experiment I, where set of symbols is a variable (taking the place of Color Combination) and level of resolution is a variable (taking the place of blink - no-blink).

5. EVALUATION: Evaluation and data analysis will be almost identical to the concepts and techniques used in Display Experiment I. Anticipated results might be that the analysis will show a difference in operator performance when using different symbology, or it might show that low resolution color is far more effective than high resolution black and white presentations. One might also expect to find interactions showing that high resolution works best for certain situations and vice versa.

C. DISPLAY EXPERIMENT III

1. EXPERIMENT TITLE: Entry Devices and Alphanumeric Displays

2. OBJECTIVE: To evaluate the usefulness of various methods for displaying alpha numeric information and the use of entry devices under a variety of situations.

3. RESOURCES REQUIRED:

a. ACCAT Testbed

b. 24 naval situation displays

c. One experimenter and 20 operators

d. Function keys and regular keyboard

e. Ability to display alphanumeric information on main display and side displays

f. Ability to record performance times

g. Ability to record subjective opinions

4. GENERAL CONTEXT

a. Concept and Need

This experiment is similar in concept design to Display Experiment I with 24 naval situations except that the variables here are:

1) Entry device

a) Keyboard

b) Function Buttons

2) Alphanumeric Display

a) On display screen anywhere

b) On split portion of screen

c) On a status display beside the main display

b. General Situation for the Experiment

The design is similar to that of Experiment I where entry devices and types of alpha numeric display are now the main variables to be investigated under different situations.

5. EVALUATION: Evaluation and data analysis will be almost identical to that of Display Experiment I. Anticipated results might show difference between:

1) Keyboard versus Function Button Entry

2) Split Screen Display versus Side Display of Alpha Numeric Information

D. DISPLAY EXPERIMENT IV

1. EXPERIMENT TITLE: Map Projection and Track History Display

2. OBJECTIVE: To evaluate mercator versus polar type map projection and different methods of displaying track history. The objective is to show the effects of these variables on transferring information to the operator.

3. RESOURCES REQUIRED:

a. ACCAT Testbed

b. 24 naval situation displays

c. One experimenter and 20 operators

d. Ability to display mercator and polar map projections

e. Ability to display track history

f. Several alternative designs for displaying track history

g. Results of previous experiments

h. Ability to record operator times and accuracy

i. Ability to record operator subjective opinions

4. GENERAL CONTEXT

a. Concept and Need

The general concept of this experiment is to evaluate the effect of different types of map projections and track history presentations on operator performance. The experiment is needed to help define which methods are best for future systems.

b. General Situation for the Experiment

The design is similar to that of Experiment I, where types of map projection and types of track history are the variables to be examined under 24 different naval situations

5. EVALUATION: Evaluation and data analysis will be almost identical to that of Display Experiment I. Anticipated results might show mercator projection to be best, along with a certain design for track history. Results might also show mercator projection to work best for certain situations but polar display to be best in other situations.

II. TECA EXPERIMENTS

EXECUTIVE SUMMARY. Experimentation on the capability and the utility of the Threat Evaluation and Countermeasures Agent (TECA) as an artificial intelligence user aid for threat evaluation at an afloat task force command and control center (TFCC) will be phased to correspond to the development of the TECA technology. The experiments described here will commence during the latter half of FY 1977 when the test resources are assembled.

Experimentation will be performed in stages ranging from simple static debug/validation experiments which concentrate on the capabilities of the TECA technology and require few resources to more extensive technical evaluations of TECA with a dynamically changing data base in a controlled environment and concluding with dynamic operational evaluations in simulated C^2 environments pitting orange and blue task forces against each other under various scenarios using WES. The staging of experimentation will allow the ACCAT team and the experimentation personnel to gain experience with the threat evaluation problem, TECA, WES, the display devices and other test resources, and it will provide adequate time for the development of WES and TECA. The results of the early phases will provide feedback that may lead to improvements in TECA and WES.

The initial experiments will consist of inputing various static threat situations concerning the status of blue and orange task forces into the ACCAT data base and observing if TECA will signal the existence of the threats, provide the threat conditions, and give recommended courses of action (COA). The test objective will be to evaluate the capabilities of TECA. Any problems detected in this validation stage with TECA should be corrected before proceeding to the technical evaluation of stage two.

The stage two experiments will be conducted in a more realistic C^2 environment with a dynamically changing data base, but they will still focus

on the capabilities of TECA. Three detailed scenarios will be used in conjunction with WES to generate threat situations and data-base updates to present a wide range of threats to be processed by TECA. The primary objective of these tests is to evaluate the capability of TECA under dynamically changing circumstances and to probe for the limits of its capabilities, e.g., what threat intensity is required to "overload" TECA, or what is the length of time from the instant threat occurs until TECA signals the threat.

In the third stage of testing, WES will again be used with orange and blue task forces which are composed as closely as possible, within the capabilities of available test resources, with equipment (sensors, weapons, etc.) and platforms that are projected for the 1980-1985 timeframe. Experimental emphasis will shift from a technical evaluation to an operational evaluation of the value of TECA to the decision maker. The objective will be to determine if TECA improves the commander's ability to make rapid and accurate decisions. Operationally realistic scenarios such as those developed at OPNAV or the Naval War College will be used in the evaluation. Each scenario will be replicated with different players with some trials having the blue forces operating with TECA and some trials without TECA. Operational measures of effectiveness, which will depend on the scenarios under play, will be used to assess the operational utility of TECA. In addition, the players and the umpire team will be subjected to post-exercise interviews and questionnaires to get subjective evaluations of the worth of TECA.

Prior to the initiation of the formal experiments, ACCAT personnel should perform pilot trials to make sure that the designed experiments test over a range of conditions that will be both realistic and informative and to head off any problems with the test resources.

1. EXPERIMENT:

a. TITLE: Stage-One TECA Experiments (Static Evaluations)

b. NUMBER: II-1

2. OBJECTIVE: Stage-one experiments will evaluate the capability of TECA to identify, in a static environment, the threats that confront an afloat task force, to describe the threat conditions, and to provide recommended courses of action. These experiments will establish the types of threats that TECA can identify and will assess the timeliness of the TECA warnings. The experiments will also serve to provide feedback through which TECA can be refined.

3. RESOURCES REQUIRED:

a. TECA and associated computer hardware

b. ACCAT data base

c. 24 threat situation data plates

1) Each data plate will be a data snapshot describing a threat situation at a given time. A given plate will include various types of threat conditions which confront the blue forces. The actual threats will correspond to the state of development of TECA; i.e., the plates for the initial tests will include only those threats that TECA will accommodate through phases 1 and 1a.

2) Each data plate will describe the situation concerning the platforms, positions, courses, speeds, sensors, weapons, fuel state, etc. in a state vector as would be generated by WES. However, WES will not be needed to generate the plates. Instead, the state vectors will be read directly into the PDP-10. Table VI shows the type of information required for one of the data plates. Figure 4 gives a graphical description of the situation and Figure 5 shows a blown-up view of the situation. A graphical description will be given for each plate for comparison with the TECA display

PLATFORM NAME	KILDEN	KYNDA	KASHIN	ECHO II	CHARLIE
TYPE	DDGS	CLGM	DDGM		
LAT	34-26	34-36	34-20	35-30	34-52
LONG	128 -30	128 -50	128 -36	129-30	129-54
COURSE	252°T	252°T	252°T	180°T	180°T
SPEED	15 KTS	15.0 KTS	15.0 KTS	5.0 KT	5.0 KT
GAL. OF FUEL	10,000	10,000	10,000	100,000	100,000
CF FUEL CONSUMP. RATE	50	50	50	1	1
SURF. MISSILE RANGE	20	SSN-3 30 NO DATA LINK	20	30 NO DATA LINK	30
GUN RANGE	8.5	8.5	8.5		
WPN1	SSN-11	SSN-3	SSN-11	SSN-3	SSN-7
WPN2	GUN-76MM	GUN-76MM	76MM		
			NOTE: FOR BLUE & ORANGE ALL SENSORS ACTIVE FULL WEAPON LOAD		
			NOTE: NO AIRCRAFT AVAILABLE TO "LINK" SSN-3, THUS 30 NM RANGE		

PLATFORM NAME	TRUXTON	SPRUANCE	KNOX	ROBISON	LOS ANGELES
TYPE	CGN	DD	FF	DDG	
LAT	34-14	34-09	34-04	34-00	33-00
LONG	128-48	128-48	128-42	128-51	128-30
COURSE	045°T	045°T	045°T	045°T	010°T
SPEED	18 KTS	18 KTS	18 KTS	18 KTS	5 KTS
GAL. OF FUEL	12,000	12,000	100	12,000	100,000
CF	55	55	50	55	1
SURF. MISSILES	SM1 - 25M	—	SM1 - 25M	40 TARTAR 15 M	SUBROC
GUNS	9 MI	9 MI	9 MI	9 MI	

TABLE VI. STATE VECTOR FOR DATA PLATE #1 34

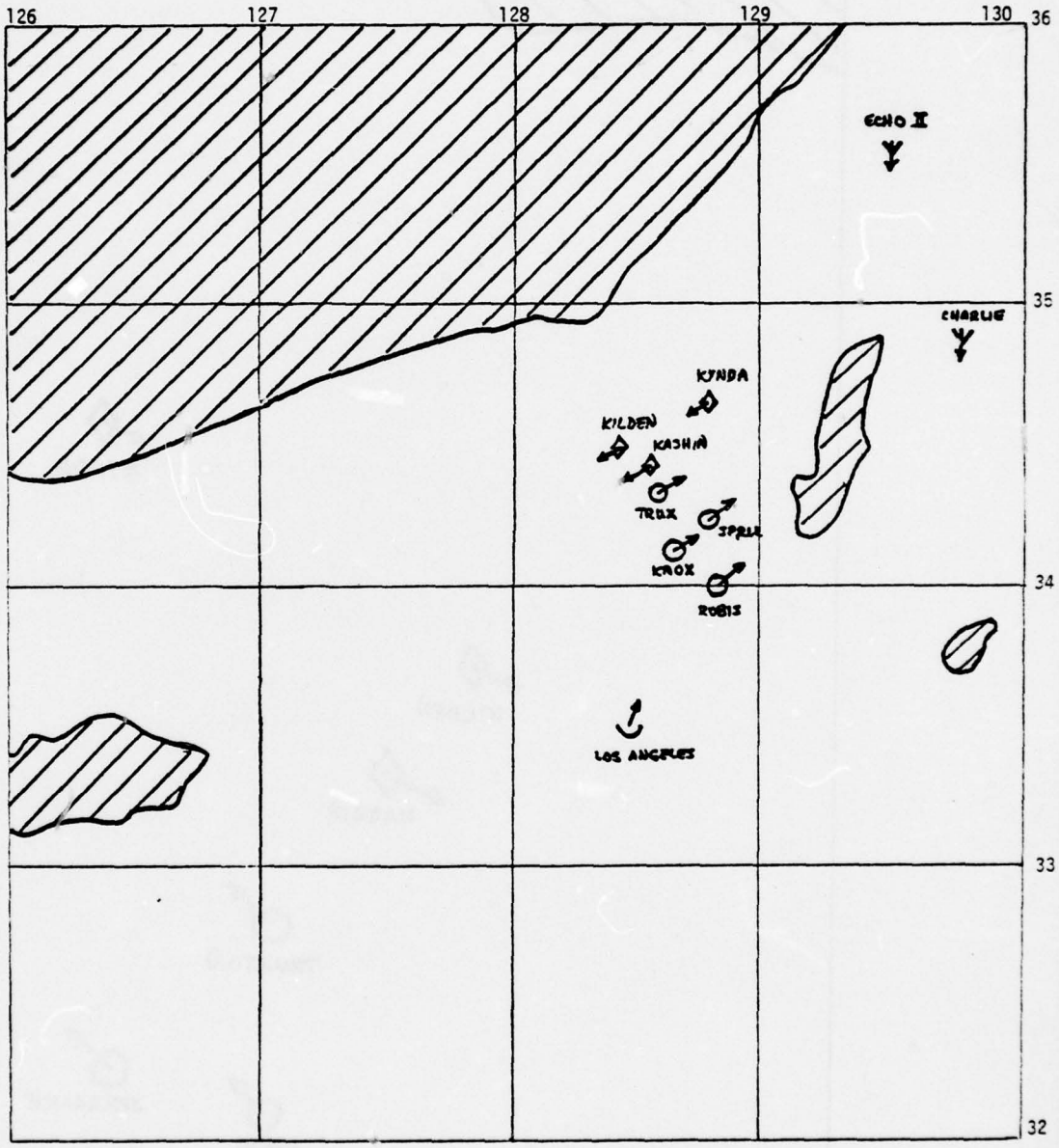


FIGURE 4. GRAPHICAL DESCRIPTION OF DATA PLATE #1

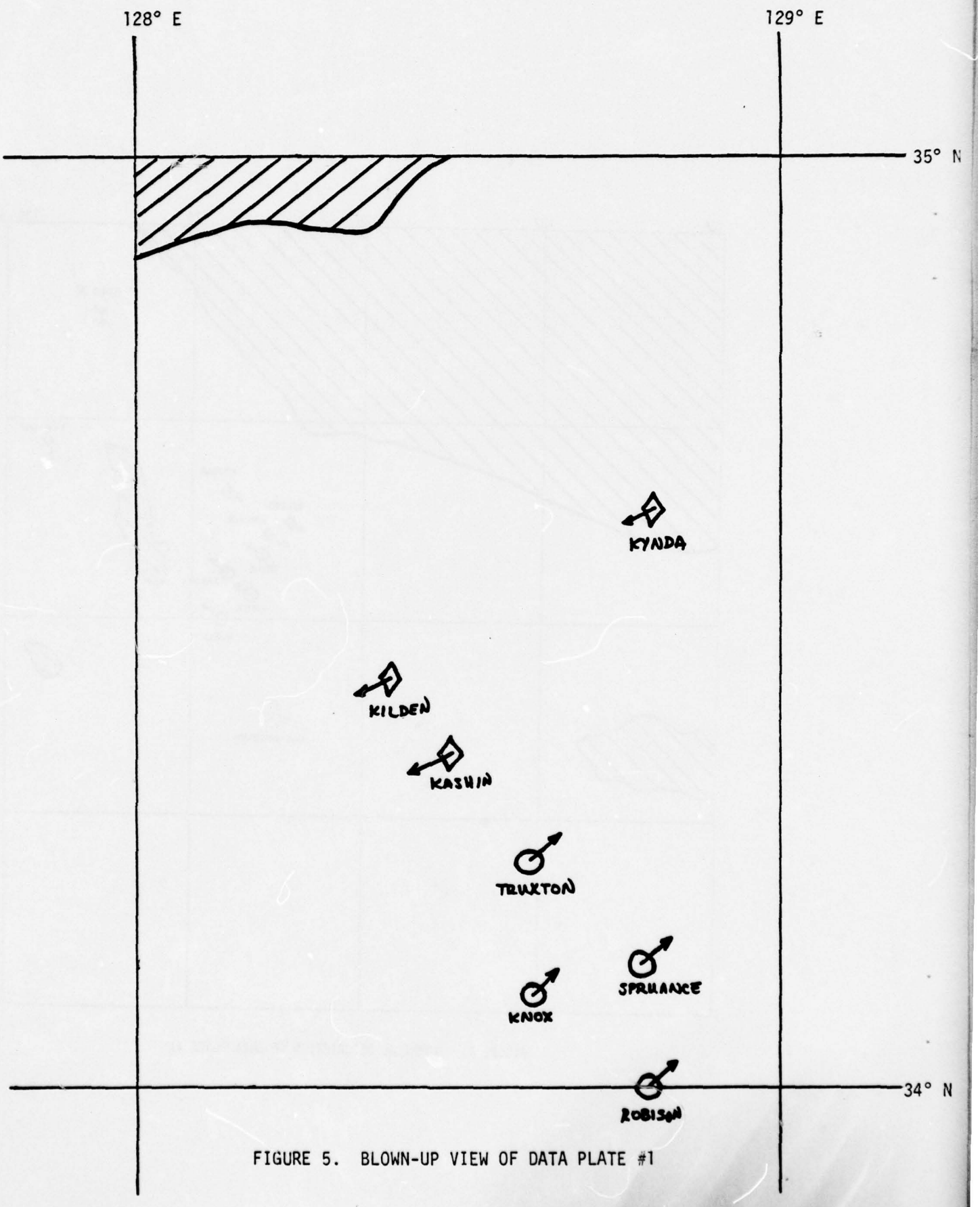


FIGURE 5. BLOWN-UP VIEW OF DATA PLATE #1

and to aid in the human evaluation of threats.

3) Personnel at the Naval Postgraduate School will aid in the development of the data plates by suggesting some of the threat situations. Primary responsibility for the development of the plates must reside in the ACCAT team since they have the hardware resources needed to generate the data plates and they are in close touch with the development of TECA.

d. Display terminal

1) Initially a monochromatic alphanumeric display terminal will be sufficient for the stage-one tests.

2) At the conclusion of Phase Ia of the TECA development, the GENISCO display system will be needed to test TECA's capability of displaying locations, identities and motion vectors of both friendly and threat platforms and their respective weapon and sensor coverage areas at the present time and at specified future times.

e. Personnel

1) The tests will require a threat assessment team to manually evaluate each data plate and identify all threats. This human threat evaluation will be done prior to the tests of TECA.

2) Two subjects will be needed during the experiments; one to enter the plates into the data base and one to monitor the TECA-driven display.

3) Post-exercise analysis personnel will be needed to compare the threat situations with the TECA descriptions of the situations and to analyze the time lags.

f. Timing device to measure the elapsed time between insertion of the data into the data base and the TECA warning signal. (A software

clock is suggested to record the times automatically. This will require a small amount of programming, but will give more reliable measurements.)

g. Software

1) Software must be written to input the data plates into the data base and to send a stimulus to notify TECA of the data-base update.

2) Software will be required to measure and record the times of the data-base updates and the times at which TECA sounds its warnings.

3) Software will be required to record all TECA-generated output messages.

4. GENERAL CONTEXT:

a. Concept and Need

TECA is being developed to interact with the command and control data base, with computational models, and with a graphics display system to support the commander's decision processes. In order to assess the value of the TECA technology we must determine its capabilities and its shortcomings. For TECA to be useful to a decision maker in a command and control environment it must be able to identify and evaluate all threats, describe the threat conditions, recommend countermeasures and display the situation in a timely manner. These experiments will assess those capabilities of TECA. Because of the evolving nature of TECA and the short time period between the delivery of the ACCAT hardware and the scheduled beginning of experimentation, the tests will begin with relatively simple experiments requiring a minimum of test resources.

b. General Situation and Scenarios

Twenty-four threat situation data plates will be developed to test the capability of TECA. The plates will be selected so as to include as many of the phase 1 threat situations, as is feasible, that TECA will

be ready to evaluate. The plates will contain multiple threats each of which would constitute a threat to some member of the friendly forces. The pilot trials should be used to explore the number and types of threat situations that should be contained in a given plate that are consistent with the state of the TECA technology.

A threat situation data plate will be selected at random, its number recorded and the data input into the ACCAT data base. The time of input should be recorded automatically by a software clock. An observer, unaware of the contents of the data plate or the input time, will monitor the display terminal receiving information from TECA. When TECA signals the existence of a threat, the time of the signal, the threat conditions and the recommended countermeasures will be recorded automatically.

The data plates need not be replicated since the outputs for a given data plate will be exactly the same from replication to replication. Thus, only 24 trials will be run. Allowing approximately five minutes per trial and a rest period half the way through the trials, the stage-one experiments will require no more than three hours to run.

5. EVALUATION:

a. Data Collection

The following data should be collected for each trial:

1) The data plate identification number

2) The time of entry of the data plate into the data base

3) The time at which TECA signals the existence of the threats

4) The time at which the TECA evaluation of the threat situation has terminated as measured when the output at the display terminal ceases

5) The threat conditions and recommended countermeasures given by

TECA (a "hard copy" printout of the displayed information). From these data (summarized in Table VIII) and the human-assessed threats performed prior to the stage-one experiments, we can determine time lags, false alarms, missed threat identifications, and TECA's threat prioritization. In addition we can evaluate the adequacy of the threat descriptions, the recommended countermeasures, and the displayed information.

INPUT TIME	PLATE #	TIME OF ALARM SIGNAL	TIME OF COMPLETION MESSAGE	THREAT CONDITION	COA
0800.05	12	0801.17	0801.55	C_{11} & C_{12} & . . . & C_{1k}	A_{11} & A_{12}
0807.13	3	0810.07	0811.32	C_{21} & C_{22} & . . . & C_{2n}	A_{21} & A_{22} & A_{23}
°				°	
°				°	
°				°	

TABLE VII: DATA SHEET FOR STAGE-ONE TECA TESTS

The human-assessed threats should be documented for comparison with the TECA output. For example, for the data plate described by Table VI and Figures 4 and 5 the threat assessment might be as described in Table VIII.

PLATFORM	THREATS
TRUXTON	1. Within effective range of SSN-11; BRG 355°T, 6 NM 2. Within effective range of SSN-11; BRG 335°T, 13 NM 3. Within effective range of SSN-3; BRG 027°T, 24 NM 4. Within effective range of 76MM Guns, BRG 335°T, 6 NM
SPRUANCE	1. Within effective range of SSN-11; BRG 317°T, 15.5 NM 2. Within effective range of SSN-3; BRG 003°T, 27 NM
KNOX	1. Within effective range of SSN-11; BRG 344°T, 17 NM 2. Warning: On present course and speed, will be within effective range of SSN-3 in 5 minutes. Threat platform now BRG 012°, 32.5 NM. Recommend come starboard 45°. 3. Warning: Will fun out of fuel in approximately 2 hours
ROBISON	1. Warning: On present course and speed, will be within effective range of SSN-3 in 40 minutes. Recommend come starboard 10°.

TABLE VIII: THREAT ASSESSMENT DATA SHEET

b. Analysis

The analysis undertaken for the stage-one experiments will be concentrated in two primary areas:

1) Comparison of the TECA output with the human assessments of the threats

2) Determination of the times required for TECA to process different types of situations

Post-test analysis will examine the data sheet to ascertain if TECA accurately evaluated all phase-one threats and gave warnings in a timely manner. The recommended countermeasures will be examined primarily for debugging purposes. The accuracy of the TECA threat descriptions and COAs will be evaluated.

c. Anticipated Results

It is anticipated that TECA will signal the existence of threats with very little time delay and that the threat conditions and countermeasures will be presented accurately. For those cases where complex multiple threat situations are present simultaneously, the tests may indicate the need for a threat prioritization routine and/or better methods for displaying threat information.

6. COMMENTS AND SPECIAL INSTRUCTIONS

Care should be taken to include in the data plates all of the types of threats that TECA has been designed to handle in realistic situations. As suggested above, these tests can be automated for the most part with no need to replicate. However, if Phase 1A of the TECA development is on schedule and the GENISCO display can be used in these experiments, we recommend that several different operators be subjected to the complete set of plates and be asked to evaluate subjectively the displayed information. Such feedback will provide useful information concerning refinements

of TECA and will provide direction for follow-on display experiments.

It is possible that the threat situation data plates could serve double duty by also being used in the display experiments. There are some differences in the input mechanisms, but perhaps software could be developed to allow the TECA plates to be used for the display experiments. The process of developing the data plates is time consuming and many of the experiments will require such static data plates. Therefore, it is important that the ACCAT team have the capability of producing such experimental resources.

B. STAGE-TWO TECA EXPERIMENTS

1. EXPERIMENT:

   a. TITLE: Dynamic Technology Assessment

   b. Number II-2

2. OBJECTIVE: To assess the technical capability of TECA actions on a dynamically changing data base generated by creating threat situations through the exercise of WES; to determine the "limits" of the capability of TECA; and to obtain subjective opinions about TECA.

3. RESOURCES REQUIRED:

   a. TECA and associated computer hardware

   b. WES

   c. ACCAT data base

   d. GENISCO display system for blue team and a display terminal for the umpire team

   e. WES generated tapes for three scenarios

   f. Four static threat situation data plates

   g. Personnel

      1) Umpire team (threat assessors)

      2) Two WES trained operators

      3) Six observers of the TECA output

      4) Test director

   h. Software to measure and record the times at which TECA messages are output and the times designated by the observers, and software to record the information displayed by TECA

   i. Tapes of the WES outputs of three scenarios involving blue and orange task forces

4. GENERAL CONTEXT:

   a. Concept and Need

      A complete technological assessment of the capability of TECA must

include an evaluation of TECA in a realistic operational environment with a dynamically changing data base. The experiments must subject TECA to a wide variety of threat conditions under varying levels of threat intensity in an effort to exercise all capabilities of TECA and to try to saturate TECA. The tests should attempt to probe the "limits" of the capability of TECA.

b. General Situation and Scenarios

Three scenarios will be generated to present a wide variety of threat conditions and varying levels of threat intensity. Two WES trained operators, representing the blue and the orange forces, will play out the war games without TECA according to the scenarios and a taped record of each game will be made. (The taped records will consist of the files of information corresponding to the periodic WES updates of state vectors of the war games.) Pilot trials should be conducted to determine the appropriate levels of threat intensity to be included in the war games. The umpire team will determine those situations and times that the blue forces were threatened by the orange forces during the war game.

Each experimental subject will go through a short TECA indoctrination session during which he will view four of the static threat situation data plates developed for the stage-one tests. The subject will also be told that he will monitor a display of the blue forces receiving threat information from TECA and that he should signal after each TECA warning when he understands the threat situation, but that he will have no control over the actions of the blue forces.

During an actual experimentation run the tape of one of the three WES games will be played back with one of the six observers monitoring the blue forces which receive threat warnings and messages from TECA.

The times at which the data base is updated, the times at which TECA signals the threats, the threat conditions as described by TECA and the recommended countermeasures should all be recorded automatically. In addition, the blue observer will signal when he feels that he comprehends the threat situation by depressing a specified key on a display keyboard. The times that the key is depressed will be recorded automatically.

The three scenarios should each have a duration of approximately two hours. Each of the six observers will observe all three scenarios. For each operator the order of the scenarios will be sequenced differently to balance out any temporal effects. The following schedule of trials should be followed:

Operator	Scenario Sequence
1	(1, 2, 3)
2	(1, 3, 2)
3	(2, 1, 3)
4	(2, 3, 1)
5	(3, 1, 2)
6	(3, 2, 1)

At the conclusion of each trial the observer will undergo a debriefing period and a questionnaire will be administered to obtain the observers' subjective assessment of TECA. Sample questions are shown in the next section.

5. EVALUATION:

a. Data Collection

The following data will be collected for each scenario:

1) The times at which that data base is updated

2) The times at which the threat conditions were signalled by TECA

3) The threat conditions as described by TECA

4) Recommended countermeasures

In addition, for each trial the time interval between each data-base update and the observer's signal of his understanding of the situation will be measured and recorded.

These data will be augmented by the threat assessments made by the umpire team and the subjective appraisals by the observers to evaluate the following information:

1) What is the delay time for TECA to acquire data and issue threat warnings?

2) What are the saturation points of TECA?

3) What are the saturation points of the subjects? How many threats can a subject comprehend?

4) What are the delay times for the subject to understand the threat situations?

5) Were there any false warnings or missed threats?

6) Were there any learning effects within observers? (Is there a sequence by observer interaction?)

7) Is there a scenario by operator interaction?

8) What are the observer's opinions about TECA?

The questionnaires given to each operator at the conclusion of each trial will consist of a few questions to identify the subject, determine his prior C^2 experience, and to identify the scenario just completed. These boilerplate questions will be followed by no more than 10 questions soliciting the observer's opinions about the following:

1) The TECA signalling device

2) The TECA output method

3) The type of information given by TECA

4) The most useful information given by TECA

5) The least useful information given by TECA

6) Features that should be added to TECA

7) The value of permitting the observer to selectively suppress some of the TECA information

8) The usefulness of TECA to a decision maker

9) The type of situation where TECA is most useful

10) The adequacy of the recommended countermeasures

b. Analysis

1) Graphical and tabular displays of the data concerning time lags and summary data such as means and variances

2) Analysis of TECA processing time as a function of the number of threats, platforms, and/or types of threats

3) Analysis of observer comprehension time as a function of the number of threats, platforms, and/or types of threats

4) Effects of learning (trial number) on observer's comprehension time and/or subjective opinions

5) Effects of scenario on observer's subjective opinions about TECA

6) Comparison between the TECA threat descriptions and those by the team of threat assessors

c. Anticipated Results

We anticipate that we will not be able to saturate TECA through Phase 1a with the restrictions on update frequency and number of platforms imposed by WES. Later, as more threat evaluation tasks are given to TECA in phases 2 and 3, we may be able to generate saturation points for TECA. We anticipate that the threat warnings will be produced in a timely manner and that the observers will consider TECA to be a valuable command and control decision aid.

These experiments should provide a thorough examination of the

capabilities and limitations of TECA. In addition, they should provide an inkling of the acceptability of TECA to the user population.

6. COMMENTS AND SPECIAL INSTRUCTIONS

Tapes of the three scenarios are requested for these experiments so that a given game can be replicated exactly with zero variance for the different operators and so that the umpire team (threat assessors) can determine one time for each scenario what threats exist. Furthermore, having the games on tape permits the umpire team to run through the same game several times to thoroughly check out all threat conditions. Also, having the games on tape would allow trials to be run without requiring WES-trained operators to input the instructions required by a detailed script. In addition, with the games on tape, an observer could monitor the blue forces in one or more of the scenarios in a run without TECA and make comparisons with the exact same run with TECA. Finally, if the runs were not available on tape, very detailed scripts describing all of the actions of both the blue and the orange forces would have to be written for each of the scenarios to accomplish the objectives of this test.

C. STAGE-THREE EXPERIMENTS

1. EXPERIMENT:

a. TITLE: Operational Evaluation of TECA

b. Number II-3

2. OBJECTIVE: To determine an operational evaluation of the military utility of TECA to a decision maker at an afloat task force command and control center in a simulated C^2 environment. To evaluate whether TECA improves the decision maker's ability to make rapid and accurate decisions.

3. RESOURCES REQUIRED:

a. TECA and associated computer hardware

b. WES

c. Three display terminals (blue, orange and umpire teams)

d. ACCAT data base

e. Five operationally realistic scenarios

f. Personnel

1) Six teams to serve as blue and orange forces

2) One umpire team

3) Two WES operators

4) Test director

g. Software to measure and record the times at which TECA messages are output, the TECA messages, the times at which instructions were given by the TECA-aided decision maker, and a record of the instructions.

h. Experimental Command Center

4. GENERAL CONTEXT

a. Concept and Need

An assessment of the value of TECA as a decision aid in an operational command and control environment can only be made by evaluating whether TECA enables a decision maker to better understand the threat situation which faces his forces and to make better and more timely decisions. These experiments will pit two opposing decision makers in a simulated C^2 environment and seek to measure the value of TECA by comparing the decision makers' performances operating with and without TECA.

b. General Situation and Scenarios

The operational evaluations of TECA will require five operationally realistic scenarios consiting of blue and orange task forces like those projected for the 1980-1985 time frame. The scenarios and the initial conditions should be selected so that a complete game can be played out in approximately three hours. The scenarios will be acted out with each team given the flexibility of exercising complete control over

their forces as long as they do not countermand their ordered missions. (The actual commands will be entered into WES by the two WES operators.) For each scenario, an operational MOE or multiple operational MOEs will be selected and used to assess the operational utility of TECA.

Because of the free-play flexibility given to the blue- and orange-force commanders, the war games and the resulting outcomes will likely vary significantly from trial to trial. Consequently, each scenario will be replicated three times with the blue forces playing one game without TECA to establish a baseline for comparison and two games with TECA. The tests will be set up so that each team will play each of the other five teams exactly one time and each time will play some of its games with TECA and some of its games without TECA. The trial matrix should be as shown in Table IX. The first entry of the pair corresponds to the orange team and the second entry of the pair corresponds to the blue team. The "T" following the blue team indicates that the blue team for that game will be aided by TECA. The restrictions on randomization imposed by this experimental design have been incorporated in an effort to balance out the learning and the team effects. All replications of scenario one should be performed, then the three replications of scenario two, etc., until all fifteen games have been played.

Replication \ Scenario	1	2	3	4	5
1	(1, 2T)	(4, 6)	(5, 1T)	(3, 5T)	(2, 3)
2	(5, 6T)	(1, 3T)	(4, 2)	(2, 6)	(1, 6T)
3	(3, 4)	(2, 5T)	(6, 3T)	(4, 1T)	(5, 4T)

TABLE IX. TRIAL MATRIX FOR STAGE-THREE EXPERIMENTS

A brief indoctrination period lasting approximately thirty minutes should precede each team's first trial. In this indoctrination period each team will be instructed as to what it will be required to do during the tests and a demonstration will be given of WES and TECA. In addition, each team will be instructed prior to each trial about the scenario, their missions, and the measures of effectiveness.

5. EVALUATION:

a. Data Collection

The following data will be collected for each trial:

1) The times at which the data base is updated

2) The times at which the threat conditions were signalled by TECA

3) The threat conditions as described by TECA

4) Recommended countermeasures

5) Actions taken by the decision maker

6) The times at which the decision maker's instructions were input to WES

7) The data required to determine the selected operational measures of effectiveness

These data will be augmented by subjective evaluations by the umpire team and by the observers. Each team playing with TECA will fill out a questionnaire like that used in stage two.

b. Analysis

1) Data summaries will be made for each scenario and tabular displays will be made of the MOEs

2) A correlation analysis will be made of the time sequence of TECA warnings with the time sequence of actions taken by the decision makers to try to determine how much the commanders utilize TECA and what information is most useful to them

3) A comparison will be made of the game outcomes as reflected by the MOEs using analysis of variance to see if there is any significant difference due to TECA.

4) An analysis of variance will also be performed to test if there were significant interactions between scenarios and TECA.

5) A comparison will be made of the game outcomes and the subjective appraisals of TECA to see if there was a relationship between how well the blue forces performed and how much they liked TECA.

c. Anticipated Results

It is anticipated that there will be large variance in the overall end-of-game operational measures of effectiveness. There will perhaps be so much "noise" in the MOEs that any signal due to TECA will not be discernible. Thus, the subjective evaluations of the teams will probably be very important. Their evaluations may provide the only discrimination in the test, for there may be no quantitative basis for discrimination as to the value of TECA.

The correlation analysis of the decision maker's actions and the TECA threat warnings should reveal the type of information that is most useful to the decision makers and how much a decision maker may grow to rely on TECA.

If the statistical noise due to team differences, learning, randomness, etc. is not so great, it is anticipated that the analysis of variance will reveal that there are significant differences due to TECA and the scenario, and that there are significant TECA by scenario interactions.

Overall, these experiments should yield important information about the acceptability of TECA to the decision maker and its utility to him.

6. COMMENTS AND SPECIAL INSTRUCTIONS

The usefulness of TECA to a decision maker may depend strongly on the manner

in which TECA outputs are displayed to the blue forces. This might be especially true in intense conflict situations if a large queue of threat conditions builds up at a given time. Thus, some experimentation in the display area should precede stage three of the TECA experiments to determine a good display candidate to be used in these trials. The blue forces display terminal should be the GENISCO display. The display system is an integral part of TECA so we should evaluate TECA with the best display system available.

Much of the data collected in these experiments can be used to evaluate the capabilities of the TECA technology as was done in stages one and two. Unless some specific problems with TECA are indicated in the trials, the type of analysis done in stages one and two will not be repeated here. The data will, however, be available for analysis. A tape record should be made of the inputs and outputs of each game so that evaluation personnel can reproduce a given trial and conduct an autopsy of what happened at any point in the game. With a tape record a lot of potentially valuable "what if" types of analyses could be conducted.

Much of the success of the experiments will depend on the scenarios, the starting conditions and the measures of effectiveness that will be used. The scenarios used should be operationally realistic, and the starting conditions and measures of effectiveness must be selected to exercise a wide range of the TECA capabilities and to yield useful information. The scenarios required for the TECA experiments should not be developed just for the purpose of exercising TECA. Instead, the blue and orange task forces should reflect as closely as possible, within the constraints of the resources, the compositions of those forces projected for the 1980-1985 time frame and the types of missions that would be probable. We

believe that the ACCAT team will have to conduct some preliminary pilot trials to select appropriate scenarios, starting conditions and measures of effectiveness Scenarios are required for just about all of the ACCAT experiments and their development could easily become a bottleneck to experimentation time schedules. It is important that the ACCAT team acquire resources for developing scenarios for testing the command-and-control technologies.

Finally, the subjects used in the experiments will be very important. For scientific experimentation purposes we need subjects who have command and control experience as decision makers. This would probably require Captains or Admirals to serve as subjects. One of the most common criticisms of experiments which attempt to determine the operational utility of a system is that the "operators" did not reflect the potential user population. This is especially a subject of criticism when the evaluations depend heavily on subjective appraisals. We realize the difficulties in getting the type of person actually needed for these experiments but we suggest that such an effort be made. The point at issue here is whether these tests are really to be viewed as scientific experiments or demonstrations. If they are to be considered as experiments, then we need operational realism so that our conclusions will be credible.

III. LADDER EXPERIMENT

EXECUTIVE SUMMARY. A static "discrete slide" of data contained in Blue File will be accessed using INLAND with its natural language input and IDA with its simple query language. The objectives are to assess training/learning requirements for operators with the two options, as well as to compare their times to recover the information requested, compare the respective error rates in formulation of query messages and to obtain subjective evaluations of the two languages.

For follow-on evaluations it is desirable to use a dynamic data base situation, with a free play "scenario" involving operators interacting with decision makers and the query systems, under varying degrees of stress and query traffic. It would be desirable, therefore, to develop a method of making WES output accessable to LADDER through the Blue File; we recommend this task be undertaken by ACCAT as a high priority goal.

1. EXPERIMENT:

   a. TITLE: LADDER Experiment to Evaluate Relative Performances with INLAND and IDA

   b. Number III-1

2. OBJECTIVE: To assess operators' ability to access a limited data base using two levels of query language; to evaluate operators' ease of learning with the two levels; to determine effects of differences in query flexibility at the two levels; and to obtain operators' subjective opinions about the two approaches to accessing the Blue File data base.

3. RESOURCES REQUIRED:

   a. Computer for use with LADDER software

   b. Display and Keyboard input device

   c. LADDER software, together with software required to measure and record times between events, to allow keyboard input, to display

information (operator instructions, questionnaire, etc.) and to record operator responses.

d. A "Blue File"-like data base containing information to be accessed

e. Two query lists, with programmed entry times

f. 8 Operators (needn't be familiar with LADDER)

g. Test director, LADDER language instructor

h. Instruction and training materials for teaching operators the IDA query language and the LADDER natural language

i. Questionnaire for presentation on display

j. Typing test, to be administered through display, with automatic recording of completion time

k. Operator briefing and prompting information for "directing" trials

4. GENERAL CONTEXT:

a. Concept and Need

The need for improvement in the man/machine interface has led to development of LADDER for use in querying certain data bases. It is not known, however, what level of query language might provide the best access to data bases for C^2 situations. In this experiment, two levels of query language, both using the same data base, are assessed and compared.

b. General Situation and Scenario

Operators are trained to use the IDA and INLAND query languages. Times required for each operator to become proficient with each language are recorded, for subsequent analysis. A training "instructor", together with training materials developed by ACCAT personnel are to be used; possibly all training can be accomplished with an interactive display package. Care should be taken to train half of the operators first with IDA, the other half with INLAND

first, since there is probably carryover from one area to the other, and we wish to evaluate the differences in training requirements with the two languages.

Operators receive displayed requests for information with requests added to the operator's "List" or queue in accordance with a pre-selected schedule. Two such programmed schedules are required, say program "A" and program "B". The operator is instructed to use the first-in-first-out discipline in processing requests. The programmed schedule of request arrivals is designed so as to provide varying load (traffic volume) and difficulty (translation of displayed "verbal" request to INLAND or IDA query form). This is done so as to enable evaluation of the operator's "learning" and fatigue effects, the operator's responses to busy periods, and the operator's request handling capacities. A number of operators (eight is suggested) would each receive two such trials, one with each query language , in accordance with the schedule shown below. The operators will be trained so they are competent with both the INLAND interface method and IDA. The trials should last roughly two hours, including an initial 15 minute "calibration" period and a final 15 minute "debriefing" period. Each operator remains at his task until he has served all the programmed requests for the trial. The total time required for a trial would vary due to differences in operators and query languages.

The calibration period at the beginning of a trial is used to bring the operator "up to speed" with the particular query language to be used in the trial, and to administer a short typing task. The typing speed and accuracy of the operator will be measured by displaying a short passage and asking the operator to key it into his display keyboard. The operator will be required to edit his

response until it contains no errors; his score is total time required. This information will later be used to determine its degree of correlation with operator performance using these query methods. The typing test will be administered only at the beginning of the first trial encountered by each operator. The calibration period will begin with a displayed briefing of the experiment, the operator's role, how long it will take, and instructions for performing the trial.

The debriefing period will be used to administer a questionnaire to each operator with the display, following the operator's second (and final) trial. The questionnaire is designed to determine the operator's subjective assessment of each query method. A sample questionnaire is shown below.

The preprogrammed schedules, A and B, should have about the same length, frequency of arrivals profile and difficulty of requests mix. The requests should all be processable using IDA. The mix of requests should "span" as much as possible this set of feasible queries. It is possible a trial and error pilot program will be required before reasonable programs can be designed, so that request arrival rates and request difficulties are not too easy nor too difficult.

A schedule of the following form should be used in conducting the experiments ("a" denotes operator's first trial, "b" his second trial; numbers in table are operator numbers).

		Program A	Program B
Language:	IDA	1a, 3b, 5a, 7b	2b, 4a, 6b, 8a
	INLAND	2a, 4b, 6a, 8b	1b, 3a, 5b, 7a

Thus, for example, operator 3 first has a trial with INLAND using program B; later he has a trial with IDA using program A.

5. EVALUATION:

a. Data Collection

It is proposed to measure or record the following during each trial:

1) Time of arrival of each request
2) Time of completion of each query message
3) The query message itself (as typed by the operator)
4) Use of error recovery feature
5) Time of receipt of answer

From these data one can determine queue lengths, arrival rates of requests, times required to prepare query messages, flow rate of completed messages, error rates in message preparation, nature of errors in message preparation, each as a function of time in the trial as well as integrated into corresponding overall measures for the trial. Analyses with these measures and the calibration and debriefing data in turn can give:

1) Amount of response time attributable to query formulation
2) Paired comparisons of performances with two query languages (paired on operator "skill")
3) Time effects during the trials (learning, fatigue, etc.)
4) Saturation points of operators

5) Why unsuccessful query attempts occurred

6) Response of operators to queue length (in terms of flow rates and error rates)

7) Differences among operators

8) Effects associated with typing ability, amount of interactive terminal experience, etc.

9) Subjective opinions of operators

b. Anticipated Statistical Analysis of Data:

1) Graphical and tabular displays of data and data summaries

2) Analysis of operator differences (test of no difference, variance estimates, operator X technology interaction characteristics, maximum process rate capabilities)

3) Effects of learning, fatigue (perhaps by regression on time or request number)

4) Effects of queue length (regression of error rate on queue length, and of operator process time on queue length)

5) Correlation analysis between: prior experience and various measures of performance (such as error rate, processing time); typing ability and various measures of performance; training time required and various measures of performance

6) Paired comparisons of various measures of performance under INLAND and IDA

7) Degree of association between subjective opinion and various measures of performances

8) Analysis of training requirements for INLAND and IDA languages, together with evaluation of differences among operators

c. Anticipated Results

We expect to find INLAND easier to learn to use, but IDA to be

more flexible in the queries it can handle and to be less time consuming in total time to answer an input request for information. There will probably be large differences in operators, in terms of learning times required, and in resultant skills in using the query languages. Error rates by operators will probably increase with increasing length of the queue of requests and with fatigue of the operators. It is expected operators may prefer the INLAND language, although this is not clear.

Overall, this experiment should provide important information about man/machine interface languages and some of the important factors affecting data base access activities. In addition, the experimentation technology developed should be useful for future experimentation at ACCAT.

6. COMMENTS AND SPECIAL INSTRUCTIONS

Early attention should be given to the problem of training operators in the two languages, and in collecting useful information about training requirement differences, if any. This will require development of a training procedure and training materials, possibly for use with an interactive terminal setup.

It is anticipated the entire test sequence for each operator can be largely automated through use of the terminal, with keyboard input and display output. This would appear to be an efficient approach, well worth the investment in development of the required software resources. It is assumed these resources will be developed by the ACCAT team.

Analysis of the data requires recorded information coded and correlated by operator, with all pertinent information available within each operators' "record". This data must be available in a form readable by the IBM system at NPS. At least 30 days are required for the analyses proposed.

A sample questionnaire for administration to operators after their second (and final) trial is shown below.

SAMPLE QUESTIONNAIRE FOR VALIDATION PHASE

1. What is your name?
2. What is your rank?
3. What is your specialty?
4. How many years have you been involved in command control operations?
5. How long ago (years) did you first use a data base query language?
6. Which query language did you like best?
   A. INLAND
   B. IDA
7. Which query language do you feel is most reliable?
   A. INLAND
   B. IDA
8. Which query language do you think would be best overall for command control use?
   A. INLAND
   B. IDA.
9. Which query language do you feel is easiest to learn?
   A. INLAND
   B. IDA

THANK YOU FOR YOUR PARTICIPATION IN THIS EXPERIMENT

END OF TRIAL

IV. RAND TERMINAL AGENTS EXPERIMENTS

EXECUTIVE SUMMARY. Five "terminal agents" (TA's) developed by Rand are available for initial ACCAT evaluation: RITA (applied to data-base querying), NED (2-dimensional text editor), VT (virtual terminal facility), REMIND (alarm clock) and MS (message system). These TA's enjoy a symbiotic relationship in a communications/message/remote data accessing environment. Two experiments are proposed, one with a static data base and simulated message source and sink, the other with a dynamic data base driven by a WES game. Both experiments are designed to allow assessment of the utilities of the TA's, in certain specified mixes, under simulated high-intensity situations requiring messages (sending and receiving), data base querys, and evaluations of situations by operators. The training required for operators to learn to effectively use the TA's, as well as generate RITA rules for specified hybred functions will be assessed.

1. EXPERIMENT:

   a. TITLE: Rand Terminal Agents Evaluation with Static Data Base

   b. Number IV-1

2. OBJECTIVE: To evaluate the utility of five Rand Terminal Agents, in various mixes, to the C^2 operator and decision maker; to evaluate training requirements for preparing an operator to effectively use the Terminal Agents; to assess the operators use of and interaction with the Terminal Agents in a simulated C^2 environment.

3. RESOURCES REQUIRED:

   a. Computer for use with Terminal Agents software

   b. RITA rules for application to data base querying

   c. 2 displays with one capable of use with the Rand Terminal Agents

   d. Terminal Agents software

   e. Software for use in managing experimentation trials, including display presentations of operator instruction, questionnaires, messages,

requests for data, etc.; keyboard entry of information (such as interaction with displays, message formulation, operator responses to questions, etc.)

f. Other input and output hardware and software as appropriate

g. Umpire for simulating message source and message receiver, and to assess operators' accuracy of evaluations

h. Appropriate data base with sufficient information for reasonable exercise of RITA query system in simulated C^2 environment

i. 3 preprogrammed request lists for information for display to operators

j. Software clock to measure and record times of occurrence of certain events

k. Training personnel and materials for training operators to use Terminal Agents

l. 6 operators

m. Test director

4. GENERAL CONTEXT:

a. Concept and Need

Various user aids are being developed to assist in the C^2 man-machine interface. One group of such aids, the "Terminal Agents" group developed by RAND, is ready for testing at ACCAT. These aids are intended to assist operators and decision makers in the C^2 arena. It is important to assess the utility of these aids in various situations, and compare their usefulness with other possible options.

This experiment involves first "training" several operators to use the Terminal Agents in order to allow evaluation of training requirements. Operators will then engage in simulated C^2 situations requiring data base querying, message receiving and sending, and evaluations. Comparisons of operator performance characteristics with several mixes

of Terminal Agents will provide information useful for evaluating more general man-machine problems as well as assessment of the Terminal Agents themselves.

b. General Situation and Scenarios

Six operators will be trained to use the TA's. The operators need not be used simultaneously in the experiment; trials will be scheduled over a long period to accommodate other experimentation requirements, if desired. Consequently, the training should be automated as much as possible, probably through use of operator interaction on the terminal (display, using the various TA's). Written material documenting the TA's should also be available for use by the operators. While this training is rather informal, it is more than simply a "hands-on" trial and error approach, in that training times required by the various operators will be recorded. Training should be performed according to the schedule shown below. In addition, during the training phase (probably near its end) operators will be tasked to write new RITA rules for inclusion in the RITA software to perform specialized functions. Again, time required to accomplish these tasks, and the success rates and causes of failure will be recorded.

Three mixes of TA's are to be compared: 1) RITA only (for data base querying), 2) RITA plus Virtual Terminal (VT), and 3) All five Terminal Agents. Since there is probably a great deal of carryover in learning from each one of these mixes to the others, efficient investigation of training requirements dictates concentrating training on the specific agents in different orders for the various observers. We therefore recommend training operators in the following sequences:

OPERATOR	RITA	VIRTUAL TERMINAL	NED, REMIND, MS
1	1	2	3
2	1	3	2
3	2	1	3
4	2	3	1
5	3	1	2
6	3	2	1

For example, operator 3 first learns to use Virtual Terminal, then VT plus RITA, and finally VT plus RITA plus NED plus REMIND plus MS. Times required to become proficient (as judged by an umpire), together with significant problems encountered by the operators should be recorded for each operator in each configuration mix. Finally, a set of two or three situations requiring specialized RITA rules to be written, inserted into the RITA package, and exercised with the data base should be presented (one at a time) to each operator. Times required to write the rule and to successfully query the data base with the rule should be recorded by the umpire, as well as recording any significant problems encountered by operators in this process.

After completing the training phase, each operator will engage in three "trials" in the experiment, one with each TA mix. Each trial involves the operator entering a simulated ongoing high-intensity C^2 situation. The operator is presented (probably via an auxiliary display) a preprogrammed sequence of requests for information, arriving in time so as to simulate requests for information by the decision maker. The operator will be tasked to write, edit and "send" messages, and to receive and analyze messages from a simulated source (perhaps the umpire).

The operator will be asked questions (via the auxiliary terminal) initiated by the umpire at prespecified points (times) in the trial. These questions are designed to determine the operators' understanding of the situation represented in the scenario, and to simulate questions asked by the decision maker. Times required to respond to the questions, accuracy of the responses, and action taken by the operator following receipt of the question will be recorded.

The six operators will use each of these TA mixes, each under different simulated C^2 situations. Thus, 3 situations, involving different but similar programmed request lists and specified umpire messages and actions need to be prepared for the experiment. The observers should undertake the mixes and programs in the following sequence:

mix:	Program: 1	2	3
1	1, 1	2, 1	3, 2
	4, 1	5, 1	6, 2
2	3, 1	1, 2	2, 3
	6, 1	4, 2	5, 3
3	2, 2	3, 3	1, 3
	5, 2	6, 3	4, 3

NOTE: Numbers in table are in the form "observer number, trial sequence number". Thus, observer 5 first uses mix 1 with program 2, next he uses mix 3 with program 1 and finally he uses mix 2 with program 3. This design constitutes a replicated latin square design.

The duration of each trial should be about one hour. No observer should have more than two trials in one day (it would be preferable to have only one trial per day, if feasible). The total duration of trial time should thus be no more than about 20 hours; training time should not be more than about five to ten hours per operator, preferably received over a period of several days.

5. EVALUATION:

a. Data Collection

Data from the training period and trial periods for each observer are to be stored in a single record, together with identification codes, trial conditions, etc., necessary for analysis. These records should be on cards or tape readable by an IBM computer system. The following data are to be obtained and recorded for each observer:

1) Training times (one for each TA group)

2) Times to write RITA rules and to access data with them;

for each trial:

3) Times of each request arrival

4) Times and nature of each message arrival

5) Time and content of each query for the data base

6) Time of each response to data base query, and accuracy of response

7) Times and texts of messages "sent" by observer

8) Umpires judgements of accuracy of analyses of situations by observer

9) Time to complete typing task (at beginning of observers' first trial)

10) Data from questionnaires administered to observer and umpire (at end of observers' third trial)

b. Analysis

Analyses of variance with time and accuracy data from the trials will be used to determine whether there were significant differences. Correlation analysis and regression will be used to associate performance

characteristics with observer and trial or TA mix characteristics. Observer and umpire subjective evaluations will be summarized and presented. Observer differences will be analyzed.

c. Anticipated Results

There will be large variance among observers. Even so, differences in performance with the various TA mixes should be discernable. Training requirements should turn out to be easily accomplished, with a day or so of "hands-on" experience, although writing successful RITA rules may turn out to be difficult for most observers. It is expected data base querying, message formulation, and analyses of the "situation" should all become easier for observers with more TA's, although large variance among observers may mask most statistical significance among them.

6. COMMENTS AND SPECIAL INSTRUCTIONS

Since there is probably a large association between some of the variables in this experiment and observers' typing facility, a typing test should be administered prior to the observers' first trial. Even if such association were not of interest to ACCAT, it would be important to make these determinations so their effects could be used to adjust for this potential source of variance among observers. In addition, correlations of various "operational" characteristics with this measure might be of interest in connection with determining future C^2 staffing requirements. The test described in experiment III-1 can be used for this purpose.

After each trial, and at each milestone in the training phase, the umpires' and observers' opinions should be collected and recorded. At the end of each observers third trial, a questionnaire should be administered to the observer and umpire via a terminal and keyboard recording system, as described for experiment I. The sample questionnaire displayed in

experiment III-1 could be modified in obvious ways to make it suitable for the present experiment.

RAND TA EXPERIMENTS (CONTINUED)

1. EXPERIMENT:

   a. TITLE: Rand Terminal Agents Evaluation with Dynamic Data Base

   b. Number IV-2

2. OBJECTIVE: To evaluate the utility of five Rand Terminal Agents to the C^2 operator and decision maker; to assess the operators' and decision makers' use of and interaction with the Terminal Agents in a simulated dynamic C^2 environment.

3. RESOURCES REQUIRED:

   a. Computer for use with Terminal Agents software

   b. RITA rules for application to data base querying

   c. Displays and software for playing WES game

   d. Display for use with Rand Terminal Agents

   e. Terminal Agents software

   f. Software for use in managing experimentation trials, including display of questionnaires and messages, and recording data obtained in experiment (such as times of certain events, message traffic, operator responses to messages, etc.)

   g. Other input and output hardware and software as appropriate

   h. Software clock to measure and record times of occurrence of certain events

   i. Two WES scenarios

   j. Four operator-decision maker teams; one umpire

   k. Test director

4. GENERAL CONTEXT:

   a. Concept and Need

      The most significant impacts of making TA's available in the command center may become most evident in a dynamic, war game situation. In

this experiment, operator/decision maker teams play WES games with and without the TA's, in simulated stressful, high intensity situations. In order to generate a situation requiring a high density of message traffic and data base querying, it may be useful to use one game scenario which calls for interrupting the TA side (during its play) with a high-priority message from the commander-in-chief changing the teams' mission. This might simulate a miaguez rescue mission situation, for example, with message traffic to and from the commander-in-chief (perhaps played by the umpire) directing the new operation.

This experiment is designed to assess the relative value of the TA's in the above described situations, concentrating on the increased ability of the team equipped with TA's to handle the man-machine interface problems so generated.

b. *General Situation and Scenarios*

Four decision maker/observer teams play two WES games, two with and two without TA's, in accordance with the following schedule:

		With TA's	Without TA's
Scenario:	1	team 1 (blue) team 3 (orange)	team 2 (orange) team 4 (blue)
	2	team 2 (blue) team 4 (orange)	team 1 (orange) team 3 (blue)

The scenarios should present a wide variety of message, query and evaluation traffic. The scenarios should build up to intensities that probe the limits of the players ability to keep up, even with the TA's. Pilot trials might prove useful to determine the appropriate levels of intensity to be included in these games.

The games should be of about two hours duration, and the teams

should not play more than one game per day. Thus, about 8 to 10 hours of game time, spread over at least four days time, are required. The teams should have operators that are experienced in using the TA's.

5. EVALUATION:

a. Data Collection

The data recorded during each game will include the following:

1) Requests by the decision makers (times of occurrence and text)

2) Responses by the operators (times of occurrence and text)

3) Message traffic to and from each command center, including times messages were received; time messages were sent; times required to prepare messages, and message texts

4) Times requests for data base information are made by the decision maker, times queries are finished by operators, times data base informations are received by the decision makers.

5) Times of game control input by decision makers (for establishing a rough measure of game "intensity")

6) Times and durations of use of any of the five TA's

7) Subjective opinions of decision makers, operators and the umpire, obtained after each game. (This would probably be most efficiently done via display delivered and keyboard recorded questionnaires, as in earlier experiments.)

b. Analysis

From the recorded times listed above, flow rates of data base queries, as a function of "game time" can be determined, and similarly for message rates. An analysis of factors influencing these flow rates will be made, including

1) Whether TA's significantly increase flow rates

2) Whether game intensity is a good predictor of the flow rates

Comparison of the abilities of teams with and without TA's to respond

to intense situations will be made through the flow rates and, more importantly, through subjective evaluations of players and the umpire.

c. Anticipated Results

There will be very large variation in team performance from game to game and team to team. These sources of noise may well mask the statistical significance of any difference due to presence or absence of TA's. Some useful insights into reponses of teams to game situations, including game intensity and message and query flow rates should emerge. Subjective opinions about the usefulness of the TA's, together with the record of their actual usage (when available) should provide useful information concerning this man-machine interface technology.

6. COMMENTS AND SPECIAL INSTRUCTIONS

Time data should be recorded both in terms of real (clock) time as well as "point in the game" time, so reactions of the teams can be associated with their stimulus, if possible. The teams need only be present in pairs, to play the WES games. Thus, there is flexibility in scheduling in that the second pair of teams (Team 3 and Team 4) might play their WES games at a much later time than do teams 1 and 2. The players used should be representative of the populations of decision makers and operators encountered in the fleet. They should not previously have played the scenarios used in this experiment.

Questionnaires much like those presented with earlier experiments should be presented to the players (decision makers, operators, umpire), probably via the display/terminal approach suggested before.

All data recorded for a given play of the WES game should be collated into one file, with necessary identification codes giving the values of the factors used in that trial (i.e., player id's, game id, which team played blue side, which team had TA's, etc.). The data should be stored

on tape readable by the IBM System. At least 30 days analysis time is required, before results can be reported.

INITIAL DISTRIBUTION LIST

	No. Copies
Defense Documentation Center Cameron Station Alexandria, Virginia 22314	2
Dean of Research Code 012 Naval Postgraduate School Monterey, California 93940	1
Library (Code 0212) Naval Postgraduate School Monterey, California 93940	2
Library (Code 55) Naval Postgraduate School Monterey, California 93940	2
CDR C. A. Rose Code 1020 NOSC San Diego, California 92152	5
Professor D. R. Barr	5
Professor G. K. Poock	5
Professor F. R. Richards Department of Operations Research Naval Postgraduate School Monterey, California 93940	5
Dr. Earl D. Sacerdoti SRI 333 Ravenswood Avenue Menlo Park, California 94025	2
Dr. Bob Anderson RAND 1700 Main Street Santa Monica, California	2
CDR Floyd Hollister IPTO DARPA 1400 Wilson Boulevard Arlington, VA 22209	2